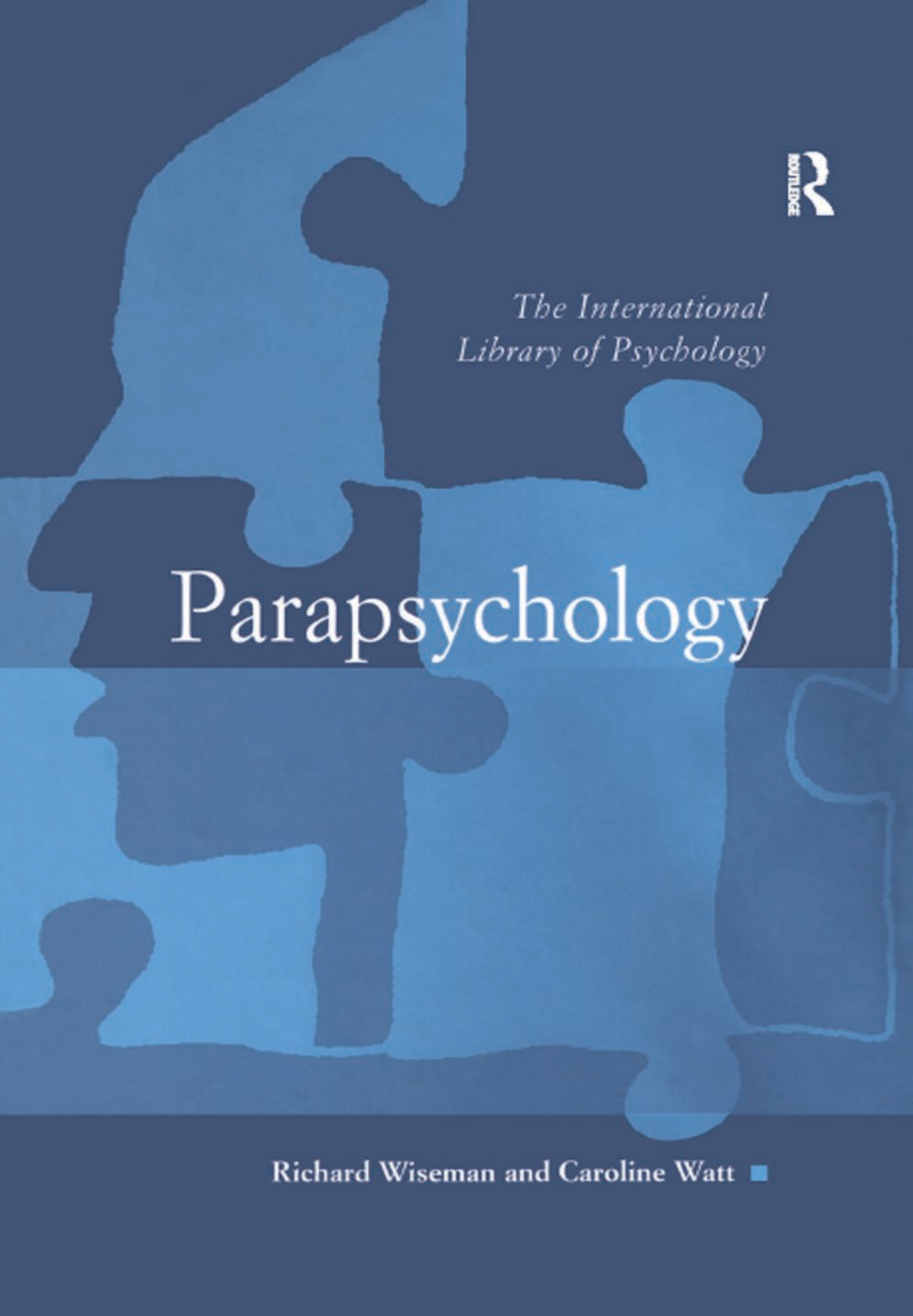
ROUTLEDGE
The International Library of Psychology
Parapsychology
Richard Wiseman and Caroline Watt

Parapsychology

The International Library of Psychology
Series Editor: David Canter

Titles in the Series

Environmental Psychology
David Canter, Terry Hartig and Mirilia Bonnes

Hypnosis
Irving Kirsch and Michael Heap

Counseling Psychology
Frederick Leong

Parapsychology
Richard Wiseman and Caroline Watt

History of Psychology Revisited
William Woodward and Sandy Lovie

Parapsychology

Edited by

Richard Wiseman and Caroline Watt

University of Hertfordshire, England and University of Edinburgh, Scotland

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2005 by Ashgate Publishing

Published 2017 by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN
711 Third Avenue, New York, NY 10017, USA

Routledge is an imprint of the Taylor & Francis Group, an informa business

Copyright © Richard Wiseman and Caroline Watt 2005. For copyright of individual articles please refer to the Acknowledgements.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Notice:
Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

British Library Cataloguing in Publication Data
Parapsychology. – (International library of psychology)
1. Parapsychology
I. Wiseman, Richard, 1966- II. Watt, Caroline
130

Library of Congress Cataloging-in-Publication Data
Parapsychology / edited by Richard Wiseman and Caroline Watt.
p. cm. — (The international library of psychology)
Includes bibliographical references.
ISBN 0-7546-2450-1 (alk. paper)
1. Parapsychology. I. Wiseman, Richard (Richard John), 1966- II. Watt, Caroline. III. Series.

BF1031.P3345 2005
133—dc22

2004054816

ISBN 13: 978-0-7546-2450-9 (hbk)

Contents

Acknowledgements ix
Series Preface xi
Introduction xiii

PART I OUTSIDE THE LAB: PARANORMAL BELIEFS AND EXPERIENCES

1 John Palmer (1979), 'A Community Mail Survey of Psychic Experiences', *Journal of the American Society for Psychical Research*, **73**, pp. 221–51. 3
2 Caroline A. Watt (1990–91), 'Psychology and Coincidences', *European Journal of Parapsychology*, **8**, pp. 66–84. 35
3 Christopher C. French (2003), 'Fantastic Memories: The Relevance of Research into Eyewitness Testimony and False Memories for Reports of Anomalous Experiences', *Journal of Consciousness Studies*, **10**, pp. 153–74. 55
4 Ian Stevenson (1993), 'Birthmarks and Birth Defects Corresponding to Wounds on Deceased Persons', *Journal of Scientific Exploration*, **7**, pp. 403–16. 77
5 Pim van Lommel, Ruud van Wees, Vincent Meyers and Ingrid Elfferich (2001), 'Near-Death Experience in Survivors of Cardiac Arrest: A Prospective Study in the Netherlands', *Lancet*, **358**, pp. 2039–45. 91
6 Richard Wiseman, Caroline Watt, Paul Stevens, Emma Greening and Ciarán O'Keeffe (2003), 'An Investigation into Alleged "Hauntings"', *British Journal of Psychology*, **94**, pp. 195–211. 99
7 Michael Daniels (2002), 'The "Brother Doli" Case: Investigation of Apparent Poltergeist-Type Manifestations in North Wales', *Journal of the Society for Psychical Research*, **66**, pp. 193–221. 117

PART II TESTING PSYCHIC CLAIMANTS

8 George P. Hansen (1985), 'A Brief Overview of Magic for Parapsychologists', *Parapsychology Review*, **16**, pp. 5–8. 149
9 John Beloff (1986), 'What is your Counter-Explanation? A Plea to Skeptics to Think Again', in P. Kurtz (ed.), *A Skeptic's Handbook of Parapsychology*, Buffalo, NY: Prometheus, pp. 359–77. 153
10 Erlendur Haraldsson and Karlis Osis (1977), 'The Appearance and Disappearance of Objects in the Presence of Sri Sathya Sai Baba', *Journal of the American Society for Psychical Research*, **71**, pp. 33–43. 173

11 Ray Hyman (1981), 'The Psychic Reading', *Annals of the New York Academy of Sciences*, **364**, pp. 169–81. 185
12 Sybo A. Schouten (1994), 'An Overview of Quantitatively Evaluated Studies with Mediums and Psychics', *Journal of the American Society for Psychical Research*, **88**, pp. 221–54. 199

PART III EXTRASENSORY PERCEPTION

13 Joseph Banks Rhine (1938), 'Experiments Bearing on the Precognition Hypothesis: I. Pre-Shuffling Card Calling', *Journal of Parapsychology*, **2**, pp. 38–54. 235
14 Charles Akers (1984), 'Methodological Criticisms of Parapsychology', in S. Krippner (ed.), *Advances in Parapsychological Research*, **4**, Jefferson, NC: McFarland, pp. 112–64. 253
15 Irvin L. Child (1985), 'Psychology and Anomalous Observations: The Question of ESP in Dreams', *American Psychologist*, **40**, pp. 1219–30. 317
16 Daryl J. Bem and Charles Honorton (1994), 'Does Psi Exist? Replicable Evidence for an Anomalous Process of Information Transfer', *Psychological Bulletin*, **115**, pp. 4–18. 329
17 Julie Milton (1999), 'Should Ganzfeld Research Continue to be Crucial in the Search for a Replicable Psi Effect? Part I. Discussion Paper and Introduction to an Electronic-Mail Discussion', *Journal of Parapsychology*, **63**, pp. 309–33. 345

PART IV PSYCHOKINESIS AND DISTANT MENTAL INFLUENCE

18 Michael Faraday (1853), 'Experimental Investigation of Table-Moving', *Athenaeum*, no. 1340, pp. 801–803. 373
19 Helmut Schmidt (1970), 'A PK Test with Electronic Equipment', *Journal of Parapsychology*, **34**, pp. 175–81. 377
20 Dean I. Radin and Roger D. Nelson (1989), 'Evidence for Consciousness-Related Anomalies in Random Physical Systems', *Foundations of Physics*, **19**, 1499–514. 385
21 William Braud, Donna Shafer and Sperry Andrews (1993), 'Further Studies of Autonomic Detection of Remote Staring: Replication, New Control Procedures, and Personality Correlates', *Journal of Parapsychology*, **57**, pp. 391–409. 401
22 Richard Wiseman and Marilyn Schlitz (1997), 'Experimenter Effects and the Remote Detection of Staring', *Journal of Parapsychology*, **61**, pp. 197–207. 421
23 John A. Astin, Elaine Harkness and Edzard Ernst (2000), 'The Efficacy of "Distant Healing": A Systematic Review of Randomized Trials', *Annals of Internal Medicine*, **132**, pp. 903–10. 433
24 James E. Alcock (2003), 'Give the Null Hypothesis a Chance: Reasons to Remain Doubtful about the Existence of Psi', *Journal of Consciousness Studies*, **10**, pp. 29–50. 441

PART V EXPERIMENTERS' PERSONAL PERSPECTIVES

25 Susan Blackmore (1987), 'The Elusive Open Mind: Ten Years of Negative Research in Parapsychology', *Skeptical Inquirer*, **11**, pp. 244–55. 465
26 Marilyn Schlitz (2001), 'Boundless Mind: Coming of Age in Parapsychology', *Journal of Parapsychology*, **65**, pp. 335–50. 477

Name Index 493

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

Acknowledgements

The editors and publishers wish to thank the following for permission to use copyright material.

American College of Physicians for the essay: John A. Astin, Elaine Harkness and Edzard Ernst (2000), 'The Efficacy of "Distant Healing": A Systematic Review of Randomized Trials', *Annals of Internal Medicine*, **132**, pp. 903–10.

American Psychological Association for the essays: Irvin L. Child (1985), 'Psychology and Anomalous Observations: The Question of ESP in Dreams', *American Psychologist*, **40**, pp. 1219–30; Daryl J. Bem and Charles Honorton (1994), 'Does Psi Exist? Replicable Evidence for an Anomalous Process of Information Transfer', *Psychological Bulletin*, **115**, pp. 4–18.

British Psychological Society for the essay: Richard Wiseman, Caroline Watt, Paul Stevens, Emma Greening and Ciarán O'Keeffe (2003), 'An Investigation into Alleged "Hauntings"', *British Journal of Psychology*, **94**, pp. 195–211. Copyright © 2003 British Psychological Society.

Copyright Clearance Center for the essay: Ray Hyman (1981), 'The Psychic Reading', *Annals of the New York Academy of Sciences*, **364**, pp. 169–81. Copyright © 1981 NYAS.

European Journal of Parapsychology for the essay: Caroline A. Watt (1990–91), 'Psychology and Coincidences', *European Journal of Parapsychology*, **8**, pp. 66–84.

Imprint Academic for the essays: Christopher C. French (2003), 'Fantastic Memories: The Relevance of Research into Eyewitness Testimony and False Memories for Reports of Anomalous Experiences', *Journal of Consciousness Studies*, **10**, pp. 153–74. Copyright © 2003 Imprint Academic, Exeter, UK; James E. Alcock (2003), 'Give the Null Hypothesis a Chance: Reasons to Remain Doubtful about the Existence of Psi', *Journal of Consciousness Studies*, **10**, pp. 29–50. Copyright © 2003 Imprint Academic, Exeter, UK.

Journal of Scientific Exploration for the essay: Ian Stevenson (1993), 'Birthmarks and Birth Defects Corresponding to Wounds on Deceased Persons', *Journal of Scientific Exploration*, **7**, pp. 403–16.

Lancet for the essay: Pim van Lommel, Ruud van Wees, Vincent Meyers and Ingrid Elfferich (2001), 'Near-Death Experience in Survivors of Cardiac Arrest: A Prospective Study in the Netherlands', *Lancet*, **358**, pp. 2039–45.

Parapsychology Foundation, Inc. for the essay: George P. Hansen (1985), 'A Brief Overview of Magic for Parapsychologists', *Parapsychology Review*, **16**, pp. 5–8.

Prometheus Books for the essay: John Beloff (1986), 'What is your Counter-Explanation? A Plea to Skeptics to Think Again', in P. Kurtz (ed.), *A Skeptic's Handbook of Parapsychology*, Buffalo, NY: Prometheus, pp. 359–77.

Rhine Research Center for the essays: Joseph Banks Rhine (1938), 'Experiments Bearing on the Precognition Hypothesis: I. Pre-Shuffling Card Calling', *Journal of Parapsychology*, **2**, pp. 38–54; Julie Milton (1999), 'Should Ganzfeld Research Continue to be Crucial in the Search for a Replicable Psi Effect? Part I. Discussion Paper and Introduction to an Electronic-Mail Discussion', *Journal of Parapsychology*, **63**, pp. 309–33; Helmut Schmidt (1970), 'A PK Test with Electronic Equipment', *Journal of Parapsychology*, **34**, pp. 175–81; William Braud, Donna Shafer and Sperry Andrews (1993), 'Further Studies of Autonomic Detection of Remote Staring: Replication, New Control Procedures, and Personality Correlates', *Journal of Parapsychology*, **57**, pp. 391–409; Richard Wiseman and Marilyn Schlitz (1997), 'Experimenter Effects and the Remote Detection of Staring', *Journal of Parapsychology*, **61**, pp. 197–207; Marilyn Schlitz (2001), 'Boundless Mind: Coming of Age in Parapsychology', *Journal of Parapsychology*, **65**, pp. 335–50.

Skeptical Inquirer for the essay: Susan Blackmore (1987), 'The Elusive Open Mind: Ten Years of Negative Research in Parapsychology', *Skeptical Inquirer*, **11**, pp. 244–55. Copyright © 1987 Skeptical Inquirer.

Springer for the essay: Dean I. Radin and Roger D. Nelson (1989), 'Evidence for Consciousness-Related Anomalies in Random Physical Systems', *Foundations of Physics*, **19**, 1499–514. Copyright © 1989 Plenum Publishing Corporation.

Society for Psychical Research for the essay: Michael Daniels (2002), 'The "Brother Doli" Case: Investigation of Apparent Poltergeist-Type Manifestations in North Wales', *Journal of the Society for Psychical Research*, **66**, pp. 193–221.

Every effort has been made to trace all the copyright holders, but if any have been inadvertently overlooked the publishers will be pleased to make the necessary arrangement at the first opportunity.

Series Preface

Psychology now touches every corner of our lives. No serious consideration of any newsworthy topic, from eating disorders to crime, from terrorism to new age beliefs, from trauma to happiness, is complete without some examination of what systematic, scientific psychology has to say on these matters. This means that psychology now runs the gamut from neuroscience to sociology, by way of medicine and anthropology, geography and molecular biology, connecting to virtually every area of scientific and professional life. This diversity produces a vibrant and rich discipline in which every area of activity finds outlets across a broad spectrum of publications.

Those who wish to gain an understanding of any area of psychology therefore either have to rely on secondary sources or, if they want to connect with the original contributions that define any domain of the discipline, must hunt through many areas of the library, often under diverse headings.

The volumes in this series obviate those difficulties by bringing together under one set of covers, carefully selected existing publications that are the definitive papers that characterize a specific topic in psychology.

The editors for each volume have been chosen because they are internationally recognized authorities. Therefore the selection of each editor, and the way in which it is organized into discrete sections, is an important statement about the field.

Each volume of the International Library of Psychology thus collects in one place the seminal and definitive journal articles that are creating current understanding of a specific aspect of present-day psychology. As a resource for study and research the volumes ensure that scholars and other professionals can gain ready access to original source material. As a statement of the essence of the topic covered they provide a benchmark for understanding and evaluating that aspect of psychology.

As this International Library emerges over the coming years it will help to specify what the nature of 21st Century psychology is and what its contribution is to the future of humanity.

DAVID CANTER
Series Editor
Professor of Psychology
University of Liverpool, UK

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

Introduction

History, Background and Terminology

Many people have experienced seemingly psychic phenomena, such as having a dream that predicts the future, seeing a ghost, or thinking of a long-lost friend and then receiving a telephone call from that person moments later. In addition, some individuals appear to possess psychic abilities, including mediums who claim to communicate with the dead, healers who seem to help cure illness, and psychics who can apparently bend keys and cutlery using just the power of their minds.

Although such allegedly psychic experiences and abilities have been reported throughout history, it is only in the last hundred years or so that researchers have carried out systematic and scientific work into these topics (for historical reviews of the area see Beloff, 1977; Hyman, 1985a). Much of the early work in this area was conducted under the auspices of one of the first organizations dedicated to the scientific study of alleged paranormal phenomena, the Society for Psychical Research (SPR). Founded in 1882 by a group of prominent academics, the majority of the SPR's initial research focused on testing individuals claiming to have strong psychic abilities, including several well-known mediums of the day. Around the turn of the last century, almost all research into alleged paranormal phenomena was conducted by individuals either working alone or on behalf of learned societies like the SPR. However, during the 1930s, Professor Joseph Banks Rhine established a parapsychology laboratory at Duke University (North Carolina, USA), and initiated the first systematic programme of university-based research into alleged psychic abilities. Rhine also pioneered a somewhat different approach to the topic, choosing to work with people who did not claim strong psychic abilities and having participants take part in easily controlled experiments, such as attempting to guess the order of a shuffled deck of cards. Since the 1930s a small number of academics have continued to conduct parapsychological research within universities throughout the world.

Most present-day researchers draw a distinction between two types of ostensible psychic ability. In extrasensory perception (ESP), a person appears to receive some information via a channel of communication not presently understood. Researchers frequently draw a distinction between three types of possible ESP phenomena: clairvoyance, in which the information received was not known to anyone else; telepathy, in which the information was known to another person; and precognition, in which the information relates to a future event. In the second type of ostensible psychic ability, psychokinesis (PK), a person appears to influence an object or their surroundings using unknown means. Researchers tend to refer to two types of alleged PK: Macro-PK, in which the apparent phenomenon is large and directly observable (for example, the levitation of an object) and micro-PK, in which small effects are produced that can only be detected via statistical analyses (for example, causing dice to roll sixes at above chance levels).

Generally speaking, researchers investigate the possible existence of ESP and PK using one of three approaches. The first is the study of various types of anomalous experience reported by

the public. These studies have involved a diverse range of methods, including, for example, attempting to identify the types of people that have such experiences and assessing the reliability of their reports. A second approach has focused on individuals who claim to be psychically gifted. These studies typically employ just one subject (the alleged psychic) and, when successful, appear to produce large and impressive effects. The third and final approach assumes that everybody possesses psychic abilities to a small degree, and usually involves carrying out laboratory-based experiments involving large numbers of individuals, none of whom claim to be especially psychic. The effects obtained in these studies are often relatively small and can only be detected by statistical analysis. All three approaches have yielded interesting and useful data, and the five Parts of this volume reflect the diversity of work undertaken in these three main areas.

In addition to employing different approaches to studying alleged psychic experiences and abilities, researchers also hold a diverse range of theoretical perspectives about such phenomena. At one end of the spectrum, some proponents argue that certain experiential and/or experimental data strongly support the existence of psychic abilities, and may believe that they understand how such abilities are best explained (for example, that they are analogous to normal sensory systems or indicative of spiritual advancement). Other researchers are less convinced by the evidence and, even if they do believe that the data suggest some form of unexplained anomaly, are uncertain about how this anomaly should best be viewed. Finally, towards the other end of the spectrum, sceptics reject the notion that there exists convincing evidence for alleged psychic abilities, and instead argue that such evidence is the result of various types of self-deception, fraud or methodological artefacts. Given these diverse viewpoints, it is perhaps unsurprising that this field has attracted a considerable amount of controversy. The essays in this volume have been chosen to provide readers with a general sense of the methods used in this research, the various viewpoints that have been advanced to account for the findings that have been obtained and the controversies generated by this work.

This Introduction is designed to help set each of the selected essays in context, and also to provide additional references for those wishing to delve deeper into the issues surrounding each of the areas covered.

Outside the Lab: Paranormal Beliefs and Experiences

As noted above, many people claim to have had experiences that they consider to be paranormal. The essays in Part I reflect researchers' attempts to understand the nature of such experiences.

The first essay (Chapter 1) describes an extensive survey, carried out by parapsychologist John Palmer, examining the incidence and correlates of allegedly psychic experiences reported by a random sample of people from North Carolina. The survey revealed that such experiences were common, were correlated with various psychological factors (for example, lucid dreaming) and often had a considerable impact on people's lives. This work also demonstrates the diversity of such experiences, with reported phenomena including alleged telepathy, poltergeist activity, apparitions, past-life memories and out-of-body experiences. Other investigators have since carried out the same type of surveys in many other countries and cultures and have obtained similar results (see, for example, Haraldsson, 1985; Roe, 1998; Schouten, 1981, 1982).

Additional work explores why people experience such unusual and seemingly paranormal phenomena. Researchers have approached this issue from a range of quite different perspectives (see, for example, Cardeña, Lynn and Krippner, 2000; Irwin, 1993; Roberts and Groome, 2001). Some researchers have argued that some anomalous experiences do not reflect the existence of genuine psychic abilities, but may instead be due to people incorrectly assigning paranormal causation to normal events (see, for example, Marks and Kammann, 1980; Shermer, 1997; Zusne and Jones, 1982). Caroline Watt's essay on coincidences (Chapter 2) represents an example of this line of research, exploring how various psychological biases may mislead people into believing that they have experienced a rare and meaningful coincidence. Similarly, Chris French's essay (Chapter 3) reviews a large body of work on the psychology of false memory, arguing that this research could help provide a normal explanation for anomalous experiences involving altered states of consciousness, such as alleged alien abductions, past-life regression and near-death experiences.

In contrast, other investigators have argued that certain anomalous experiences may provide evidence to support the existence of genuine psychic phenomena. The next two essays illustrate this approach. In the first of these, Ian Stevenson (Chapter 4) describes a series of unusual case studies that appear to support claims of reincarnation. Stevenson has gained a considerable reputation for carefully documenting cases (mainly from India and Sri Lanka) in which people allegedly remember details of past lives (see, for example, Stevenson, 1974, 1997; for a critical review of this work see Edwards, 1996) and, in this essay, he argues that certain birthmarks may be indicative of illnesses and accidents suffered by individuals in a previous life. In Chapter 5, Pim van Lommel and his colleagues present the details of a recent study into near-death experiences (NDEs). People reporting NDEs describe a remarkably similar set of phenomena, including moving through a tunnel of light, blissful feelings, life review and so on (Moody, 1975; Ring, 1980). Some researchers have suggested that these experiences could be the result of various types of hallucination (Blackmore, 1993), whereas others have argued that they may reflect some form of genuine separation between mind and body (for example, Parnia *et al.*, 2001). Van Lommel compares data from those who reported NDEs and those who did not, and argues that existing medical, pharmaceutical and psychological explanations cannot account for these experiences.

The final two essays in Part I reflect two quite different approaches to investigating very different types of alleged paranormal experiences, namely hauntings and poltergeist activity. Parapsychologists have conducted a considerable amount of work at allegedly haunted locations, examining both the reliability of eyewitness reports and whether any environmental factors (for example, air temperature, magnetic field strength and so on) are associated with such reports (see, for example, Houran and Lange, 1998; Maher and Schmeidler, 1975). The essay by Richard Wiseman and his colleagues (Chapter 6) illustrates how these types of method were used to empirically examine two well-known, and allegedly haunted, locations. Research into alleged poltergeist activity has generated a considerable amount of controversy, with some researchers arguing that the phenomena represent genuine paranormal activity (Fontana, 1991) and others that they are the result of self-deception and fraud (Randi, 1985). Mike Daniels' essay (Chapter 7) describes a recent investigation into alleged poltergeist activity in Wales, and illustrates not only the methods used in such investigations, but also the difficulties encountered when trying to reach any firm conclusion during this kind of work.

Testing Psychic Claimants

Many people claim to possess impressive psychic abilities. For example, faith healers and psychic surgeons sometimes appear able to psychically cure illness, mediums claim to contact the dead and some fortune-tellers seem to accurately predict the future. Both parapsychologists and sceptics have carried out a great deal of research into the validity of such claims, and the essays in Part II reflect the diversity of the methods and the results obtained in this work.

The opening essay (Chapter 8), written by George Hansen, notes that many investigations into individuals claiming strong psychic abilities have revealed evidence of trickery and fraud. For this reason it is vital that researchers carrying out such investigations are aware of the considerable literature in conjuring describing ways of faking such abilities. Hansen provides a brief introduction to this work and outlines its relevance for parapsychology (see also Hansen, 1990; Truzzi, 1997).

John Beloff's essay (Chapter 9) raises several important issues relating to how one assesses the investigations of alleged psychics that were carried out around the turn of the century. Beloff outlines a series of sittings that were held with a well-known Italian medium of the day named Eusapia Palladino, and outlines why he believes the events described in this material cannot have a normal explanation. Beloff's essay was first published in 1986 and, since then, the evidential status of much of the material he discusses has been the topic of fierce debate within the pages of the *Journal of the Society for Psychical Research* (see Barrington, 1992, 1993; Fontana, 1992, 1993; Polidoro and Rinaldi, 1998; Wiseman, 1992, 1993a,b,c).

The next essay, written by Erlendur Haraldsson and Karlis Osis (Chapter 10) illustrates some of the issues that arise when trying to assess an alleged psychic in a real world, and uncontrolled, situation. Haraldsson and Osis describe their encounters with Sri Sathya Sai Baba, an Indian religious leader who claims to be able to perform various miracles. Haraldsson and Osis present a detailed description of the phenomena they witnessed and their attempts to assess whether these events have a normal or paranormal explanation. Since the publication of this paper in 1977, several other researchers have written about Sai Baba's alleged abilities (see, for example, Haraldsson and Wiseman, 1995; Beyerstein, 1996).

Ray Hyman is one of the leading critics of parapsychology and has written extensively on the various methodological problems that can hinder research in this area (see, for example, Hyman 1981, 1985b, 1995). Here, his essay (Chapter 11) outlines some of the stratagems used by fortune-tellers to mislead people into believing that they are psychic (for example, the use of general statements, fishing for information and so on), and illustrates why it is highly problematic, if not impossible, to determine whether psychic readings made in a real-world setting constitute evidence for the paranormal. Many of the techniques mentioned by Hyman (such as the 'Barnum Effect', wherein people believe that very general statements, such as 'you have a great deal of untapped creative potential', are highly accurate descriptions of their personalities) have been investigated by researchers interested in persuasion and thus help illustrate the overlap between parapsychology and mainstream psychology. Researchers wishing to test psychics and mediums under controlled conditions have devised several procedures for minimizing the various types of biases and problems outlined by Hyman. The next essay (Chapter 12), written by Sybo Schouten, presents a critical review of these methods and reviews the mixed results obtained in this type of research.

Extrasensory Perception

As noted above, the first systematic programme of research into the existence of ESP was initiated by J.B. Rhine in the 1930s. Much of Rhine's work involved participants attempting to guess the order of shuffled decks of 'ESP cards' (that is, cards printed with one of five simple symbols – a circle, cross, square, star or wavy lines – on their faces). The first essay in Part III, written by Rhine in 1938, describes an initial set of precognition experiments and illustrates the type of methods involved in these early and ground-breaking studies.

Rhine's research generated a significant amount of controversy, with proponents arguing that the results supported the existence of ESP and critics claiming that the studies possessed various methodological and statistical problems (see Palmer, 1986, for a review). This, combined with the rather tedious procedures involved in the studies, and with a tendency for initially significant results to decline over time (see the review by Palmer, 1978) eventually resulted in researchers exploring other ways of running laboratory-based ESP experiments. In Chapter 14 Charles Akers presents a comprehensive and critical review of the key ESP studies that were conducted between the end of the Rhine era and the early 1980s. This review describes the many artefacts and biases that can hinder this research and then evaluates the degree to which these problems were present in a series of studies that both obtained positive results and were seen as making a significant contribution to the field.

In Chapter 15 Irvin Child reviews a series of well-known studies, conducted in the late 1960s and early 1970s at the Maimonides Medical Centre, exploring the possible existence of ESP in dreams. These studies obtained highly significant results, suggesting that the content of participants' dreams reflected randomly selected target material (for example, pictures) that were shown to them the following morning. Child also notes how much of the critical commentary attacking this work misrepresented both the methods and results of the studies.

Partly as a result of the success of the dream ESP work, many researchers focused their attention on running studies in which participants are placed into an altered state of consciousness. Much of this work involves participants undergoing the 'ganzfeld' procedure (a mild sensory deprivation procedure originally developed by perceptual psychologists to help people generate imagery) and then attempting to identify target material, such as a picture or film, being looked at by another person in a separate room.

Parapsychological research using the ganzfeld procedure has generated a considerable amount of debate, with some researchers arguing that the work represents some of the best evidence in favour of ESP (see Honorton, 1985; Honorton *et al.*, 1990; Utts, 1991), and others questioning the validity and quality of the studies (see Blackmore, 1987; Hyman, 1985b; Scott, 1986). Daryl Bem's and Charles Honorton's essay, published in 1994 and now reprinted as Chapter 16, first presents a review of the early ganzfeld studies and then describes a meta-analysis of a series of well-controlled and highly statistically significant ganzfeld studies conducted at the Psychophysical Research Laboratories. Bem's and Honorton's essay provoked a considerable amount of debate and additional meta-analyses (see Bem, 1994; Hyman, 1994; Milton and Wiseman, 1999; Storm and Ertel, 2001). Towards the end of their essay, the authors note the importance of a broad range of investigators attempting to replicate the ganzfeld ESP effect. In the following essay (Chapter 17) Julie Milton begins by picking up on this issue, arguing that the effect has declined in recent ganzfeld studies. She then discusses various reasons for this decline and outlines possible strategies for future research in this area. Milton's essay acted as

the basis for a large-scale electronic discussion about these issues, and the debate about the replicability of ganzfeld-ESP findings continues (Bem, Palmer and Broughton, 2001).

Psychokinesis and Distant Mental Influence

The first essay in Part IV is the oldest in the entire volume, and is included to reflect the type of work that helped form the foundation for more formal studies within parapsychology and psychical research. In it, the famous physicist Michael Faraday investigates 'table-turning' – a popular nineteenth-century pastime where people apparently made contact with the spirit world via the tipping and movement of a table. Faraday describes a series of experiments demonstrating that the phenomenon was not due to spirit intervention, but rather to people unconsciously pushing and pulling the table.

In addition to conducting ground-breaking research into ESP, J.B. Rhine and his colleagues at Duke University also carried out a series of laboratory-based PK studies in which participants were asked to influence mentally the roll of dice (for example, Rhine and Rhine, 1943). Some researchers argued that the resulting studies provide evidence to support the existence of PK but Girden (1962) criticized the work on several grounds, including poor methodological controls and statistical artefacts. Following only partial rebuttal of these criticisms (see Girden *et al.*, 1964; Pratt, 1963), and coupled with difficulties in replicating reported effects, researchers began replacing mechanical targets (for example, the roll of dice) with random event generators (REGs) whose output is determined by, for example, the decay of a radioactive element or from a source of electronic noise. In these studies participants are asked to attempt to bias the output of the REG in a specified direction. The essay by Helmut Schmidt (Chapter 19) describes the methodology and results of one of the first REG PK studies. This study involved participants attempting to influence the random decay of a radioactive source and obtained statistically significant results. Schmidt's studies have been the focus of a considerable amount of debate concerning the possibility of participants cheating (Hansel, 1981; Rao and Palmer, 1987) and possible REG biases in certain studies (Alcock, 1990; Hyman, 1981; Palmer, 1996; Schmidt, 1987).

The following essay (Chapter 20), by Dean Radin and Roger Nelson, was published in 1989 and presents a meta-analytic review of REG studies conducted between 1959 and 1987. The essay notes that the overall effect is statistically significant and also assesses, but rejects, the possibility that this result is due to poor methodology or a file drawer effect (that is, the selective publication of studies obtaining positive results). Subsequent experiments by an international consortium of three laboratories have raised issues about the replicability of this effect (Jahn *et al.*, 2000). Like the ganzfeld, therefore, there seems to be some inconsistency in the findings of the PK REG studies.

Other PK work has used a range of biological systems as targets. Much of this research has examined the possible influence of intentionality on human physiology and behaviour. In Chapter 21 William Braud and his colleagues illustrate the types of method used in this work and describe a study looking at one of the most common of alleged paranormal experiences – the remote detection of staring. As noted by Braud *et al.*, a large percentage of the public has had the feeling that someone is staring at them, only to turn around and discover that this is the case. Researchers have carried out several studies into this phenomenon, with mixed

results. Braud *et al.* review this research and then describe the methods and results of their own study.

The next essay (Chapter 22), by Richard Wiseman and Marilyn Schlitz, also describes a study examining the remote detection of staring, but in addition examines the possible role that the experimenter may play in determining study outcome. As noted in both this section and the previous one, much of the debate concerning the existence of psychic ability revolves around the degree to which the experimental evidence for such abilities can be replicated across several laboratories. This issue is especially problematic within parapsychology as some experimenters have a reputation for consistently achieving positive results whilst others obtain chance findings. Attempting to understand such 'experimenter effects' is therefore clearly vital to the future of the field, and the Wiseman–Schlitz study explores this issue by examining whether two studies, using the same design but carried out by different experimenters, obtain significantly different results.

A review of the small number of studies to empirically examine the possibility of 'distant healing' (including, for example, prayer and therapeutic touch) constitutes the next essay. This review, by John Astin and his colleagues, examined 23 experiments and concluded that, although the results of these studies revealed an overall effect, the poor methodological quality of the work made any clear-cut interpretation of this work problematic. Several other studies into distant healing have been conducted since this review, with mixed results (for example, Leibovici, 2001; Roberts, Ahmed and Hall, 2004).

To close the two Parts on laboratory research into ESP and PK, we have an essay by noted parapsychology critic, James Alcock (see Alcock, 1981, 1985, 1987, 1990). Alcock's essay, which forms his editorial introduction for a special issue of the *Journal of Consciousness Studies*, argues that parapsychologists are inclined to search for paranormal interpretations of their data and tend to neglect the possibility that there simply is no psi in their experiments. He gives a list of 12 'reasons to remain doubtful about the existence of psi'.

Experimenters' Personal Perspectives

The two essays in Part V are written by researchers who are both very knowledgeable about parapsychology, and have spent much of their working lives investigating the possible existence of psychic ability. However, they each present a very different perspective on what has been achieved to date and the best direction for future work in this area.

The first essay (Chapter 25) is written by sceptic Susan Blackmore (see Blackmore, 1988, 1998). Blackmore describes how she has spent over ten years examining the evidence for ESP and PK, including several studies of her own, but remains unconvinced by the evidence. She explores the problems associated with trying to prove or disprove the existence of the paranormal and concludes by suggesting that researchers would do better to focus on trying to understand people's anomalous experiences rather than discover whether such experiences reflect genuine psychic abilities.

In contrast, Marilyn Schlitz's essay formed the basis for her Presidential Address to the Parapsychological Association in 2000. Schlitz (Chapter 26) describes how she has consistently obtained evidence for psi in the experiments that she has conducted (for example, Schlitz and Honorton, 1992; Schlitz and Braud, 1997) and argues that parapsychology is alive and

well, and has the potential to inform several important academic and social issues in the near future.

Implications and the Need for Future Research

The essays in this volume were chosen to illustrate the range of topics investigated within parapsychology, the methods employed to examine these topics and the type of controversies generated by this work. Many of the essays also discuss the significant contribution that parapsychological research has already made, and has the potential to continue to make, to mainstream science. On a methodological level, such work has helped highlight several issues that are relevant to research within the social sciences, including, for example, the need to consider how experimenter effects and subtle statistical artifacts may influence the outcome of a wide range of studies. At a more theoretical level, research into the psychological mechanisms that cause people to erroneously believe that they have experienced paranormal phenomena is highly relevant to many aspects of mainstream cognitive and social psychology, such as work concerned with the psychology of deception and self-deception. In addition, evidence demonstrating the existence of ESP or PK would have profound implications for many aspects of science, including the neuropsychological modelling of brain functions and the causal nature of consciousness. The methodological and theoretical importance of these issues clearly justifies future work in this area. However, given the complexity and long-running nature of past research into these topics, it seems likely that the debate will continue for some time yet.

References

Alcock, J.E. (1981), *Parapsychology: Science or Magic?*, London: Pergamon.

Alcock, J.E. (1985), 'Parapsychology as a "Spiritual Science"', in P. Kurtz (ed.), *A Skeptic's Handbook of Parapsychology*, Buffalo, NY: Prometheus.

Alcock, J.E. (1987), 'Parapsychology: Science of the Anomalous or Search for the Soul?', *Behavioral and Brain Sciences*, **10**, pp. 553–65.

Alcock, J.E. (1990), *Science and Supernature: A Critical Appraisal of Parapsychology*, Buffalo, NY: Prometheus.

Barrington, M.R. (1992), 'Palladino and the Invisible Man Who Never Was', *Journal of the Society for Psychical Research*, **58**, pp. 324–40.

Barrington, M.R. (1993), 'Palladino, Wiseman and Barrington: Ten Brief Replies', *Journal of the Society for Psychical Research*, **59**, pp. 196–98.

Beloff, J. (1977), 'Historical Overview', in B.B. Wolman (ed.), *Handbook of Parapsychology*, New York: McFarland.

Bem, D.J. (1994), 'Response to Hyman', *Psychological Bulletin*, **115**, pp. 25–27.

Bem, D.J., Palmer, J. and Broughton, R.S. (2001), 'Updating the Ganzfeld Database: A Victim of its own Success?', *Journal of Parapsychology*, **65**, pp. 207–18.

Beyerstein, D. (1996), 'Sai Baba', in G. Stein (ed.), *The Encyclopedia of the Paranormal*, New York: Prometheus Books, pp. 653–57.

Blackmore, S.J. (1987), 'A Report of a Visit to Carl Sargent's Laboratory', *Journal of the Society for Psychical Research*, **54**, pp. 186–98.

Blackmore, S.J. (1988), 'Do We Need a New Psychical Research?', *Journal of the Society for Psychical Research*, **55**, pp. 49–59.

Blackmore, S.J. (1993), *Dying to Live: Science and the Near-death Experience*, London: Grafton.
Blackmore, S.J. (1998), 'Abduction by Aliens or Sleep Paralysis?', *Skeptical Inquirer*, **22**, pp. 23–28.
Cardeña, E., Lynn, S.J. and Krippner, S. (eds) (2000), *Varieties of Anomalous Experience: Examining the Scientific Evidence*, Washington, DC: American Psychological Association.
Edwards, P. (1996), *Reincarnation: A Critical Examination*, Amherst, NY: Prometheus Books.
Fontana, D. (1991), 'A Responsive Poltergeist: A Case from South Wales', *Journal of the Society for Psychical Research*, **57**, pp. 385–402.
Fontana, D. (1992), 'The Feilding Report and the Determined Critic', *Journal of the Society for Psychical Research*, **58**, pp. 341–50.
Fontana, D. (1993), 'Palladino (?) and Fontana: The Errors are Wiseman's Own', *Journal of the Society for Psychical Research*, **59**, pp. 198–203.
Girden, E. (1962), 'A Review of Psychokinesis (PK)', *Psychological Bulletin*, **59**, pp. 353–88.
Girden, E., Murphy, G., Beloff, J., Eisenbud, J., Flew, A., Rush, J.H., Schmeidler, G. and Thouless, R.H. (1964), 'A Discussion of Psychokinesis', *International Journal of Parapsychology*, **6**, pp. 25–137.
Hansel, C.E.M. (1981), 'A Critical Analysis of H. Schmidt's Psychokinesis Experiments', *Skeptical Inquirer*, **5**(3), pp. 26–33.
Hansen, G.P. (1990), 'Deception by Subjects in Psi Research', *Journal of the American Society for Psychical Research*, **84**, pp. 25–80.
Haraldsson, E. (1985), 'Representative National Surveys of Psychic Phenomena: Iceland, Great Britain, Sweden, USA and Gallup's Multinational Survey', *Journal of the Society for Psychical Research*, **53**, pp. 145–58.
Haraldsson, E. and Wiseman, R. (1995), 'Reactions to and an Assessment of a Videotape on Sathya Sai Baba', *Journal of the Society for Psychical Research*, **60**, pp. 203–13.
Honorton, C. (1985), 'Meta-analysis of Psi Ganzfeld Research: A Response to Hyman', *Journal of Parapsychology*, **49**, pp. 51–91.
Honorton, C., Berger, R.E., Varvoglis, M.P., Quant, M., Derr, P., Schechter, E. and Ferrari, D.C. (1990), 'Psi Communication in the Ganzfeld: Experiments with an Automated Testing System and a Comparison with a Meta-analysis of Earlier Studies', *Journal of Parapsychology*, **54**, pp. 99–139.
Houran, J. and Lange, R. (1998), 'Rationale and Application of a Multi-Energy Sensory Array in the Investigation of Haunting and Poltergeist Cases', *Journal of the Society for Psychical Research*, **62**, pp. 324–36.
Hyman, R. (1981), 'Further Comments on Schmidt's PK Experiments', *Skeptical Inquirer*, **5**, pp. 34–40.
Hyman, R. (1985a), 'A Critical Historical Overview of Parapsychology', in P. Kurtz (ed.), *A Skeptic's Handbook of Parapsychology*, Buffalo, NY: Prometheus.
Hyman, R. (1985b), 'The Ganzfeld Psi Experiment: A Critical Appraisal', *Journal of Parapsychology*, **49**, pp. 3–49.
Hyman, R. (1994), 'Anomaly or Artifact? Comments on Bem and Honorton', *Psychological Bulletin*, **115**, pp. 19–24.
Hyman, R. (1995), 'Evaluation of the Program on Anomalous Mental Phenomena', *Journal of Parapsychology*, **59**, pp. 321–51.
Irwin, H.J. (1993), 'Belief in the Paranormal: A Review of the Empirical Literature', *Journal of the American Society for Psychical Research*, **87**, pp. 1–39.
Jahn, R., Dunne, B., Bradish, G., Dobyns, U., Lettieri, A., Nelson, R., Mischo, J., Boller, E., Bösch, H., Vaitl, D., Houtkooper, J. and Walter, B. (2000), 'Mind/Machine Interaction Consortium: PortREG Replication Experiments', *Journal of Scientific Exploration*, **14**, pp. 499–555.
Liebovici, L. (2001), 'Effects of Remote, Retroactive Intercessory Prayer on Outcomes in Patients with Bloodstream Infection: Randomised Controlled Trial', *British Medical Journal*, **323**, pp. 1450–51.
Maher, M.C. and Schmeidler, G.R. (1975), 'Quantitative Investigation of a Recurrent Apparition', *Journal of the American Society for Psychical Research*, **69**, pp. 341–51.
Marks, D.F. and Kammann, R. (1980), *The Psychology of the Psychic*, Buffalo, NY: Prometheus Books.
Milton, J. and Wiseman, R. (1999), 'Does Psi Exist? Lack of Replication of an Anomalous Process of Information Transfer', *Psychological Bulletin*, **125**, pp. 387–91.
Moody, R.A. (1975), *Life after Life*, Covington, GA: Mockingbird Books.

Palmer, J. (1978), 'Extrasensory Perception: Research Findings', in S. Krippner (ed.), *Advances in Parapsychological Research, Volume 2: Extrasensory Perception*, New York: Plenum Press.
Palmer, J. (1986), 'The ESP Controversy', in H.L. Edge, R.L. Morris, J. Palmer and J.H. Rush, *Foundations of Parapsychology: Exploring the Boundaries of Human Capability*, Boston, MA: Routledge & Kegan Paul.
Palmer, J. (1996), 'Evaluation of a Conventional Interpretation of Helmut Schmidt's Automated Precognition Experiments', *Journal of Parapsychology*, **60**, pp. 149–70.
Parnia, S., Waller, D.G., Yeates, R. and Fenwick, P. (2001), 'A Qualitative and Quantitative Study of the Incidence, Features and Aetiology of Near Death Experiences in Cardiac Arrest Survivors', *Resuscitation*, **48**, pp. 149–56.
Polidoro, M. and Rinaldi, G.M. (1998), 'Eusapia's Sapient Foot: A New Consideration of the Feilding Report', *Journal of the Society for Psychical Research*, **62**, pp. 242–56.
Pratt, J.G. (1963), 'The Girden-Murphy Papers on PK', *Journal of Parapsychology*, **27**, pp. 199–209.
Randi, J. (1985), 'The Columbus Poltergeist Case', *Skeptical Inquirer*, Spring, pp. 221–35.
Rao, K.R. and Palmer, J. (1987), 'The Anomaly Called Psi: Recent Research and Criticism', *Behavioral and Brain Sciences*, **10**, pp. 539–643.
Rhine, L.E. and Rhine, J.B. (1943), 'The Psychokinetic Effect: I. The First Experiment', *Journal of Parapsychology*, **7**, pp. 20–43.
Ring, K. (1980), *Life at Death: A Scientific Investigation of Near Death Experience*, New York: Coward, McCann and Geoghegan.
Roberts, L., Ahmed, I. and Hall, S. (2004), 'Intercessory Prayer for the Alleviation of Ill Health (Cochrane Review)', in *The Cochrane Library, Issue 1*, Chichester: John Wiley & Sons Ltd.
Roberts, R. and Groome, D. (eds) (2001), *Parapsychology: The Science of Unusual Experience*, London: Arnold.
Roe, C.A. (1998), 'Belief in the Paranormal and Attendance at Psychic Readings', *Journal of the American Society for Psychical Research*, **91**, pp. 25–51.
Schlitz, M.J. and Braud, W. (1997), 'Distant Intentionality and Healing: Assessing the Evidence', *Alternative Therapies*, **3**, pp. 62–73.
Schlitz, M.J. and Honorton, C. (1992), 'Ganzfeld Psi Performance with an Artistically Gifted Population', *Journal of the American Society for Psychical Research*, **86**, pp. 83–98.
Schmidt, H. (1987), 'Alcock's Critique of Schmidt's Experiments', *Behavioral and Brain Sciences*, **10**, p. 609.
Schouten, S.A. (1981), 'Analysing Spontaneous Cases: A Replication Based on the Sannwald Collection', *European Journal of Parapsychology*, **4**, pp. 9–48.
Schouten, S.A. (1982), 'Analysing Spontaneous Cases: A Replication Based on the Rhine Collection', *European Journal of Parapsychology*, **4**, pp. 113–58.
Scott, C. (1986), 'Comment on the Hyman-Honorton Debate', *Journal of Parapsychology*, **50**, pp. 349–50.
Shermer, M. (1997), *Why People Believe Weird Things: Pseudoscience, Superstition, and Other Confusions of our Time*, New York: W.H. Freeman and Company.
Stevenson, I. (1974), *Twenty Cases Suggestive of Reincarnation*, Charlottesville: University Press of Virginia.
Stevenson, I. (1997), *Where Reincarnation and Biology Intersect*, Westport, CT: Praeger Publishers.
Storm, L. and Ertel, S. (2001), 'Does Psi Exist? Comments on Milton and Wiseman's (1999) Meta-analysis of Ganzfeld Research', *Psychological Bulletin*, **127**, pp. 424–33.
Truzzi, M. (1997), 'Reflections on the Sociology and Social Psychology of Conjurors and their Relations with Psychical Research', in S. Krippner (ed.), *Advances in Parapsychological Research*, Vol. 8, Jefferson, NC: McFarland, pp. 221–71.
Utts, J. (1991), 'Replication and Meta-analysis in Parapsychology', *Statistical Science*, **6**, pp. 363–78.
Wiseman, R. (1992), 'The Feilding Report: A Reconsideration', *Journal of the Society for Psychical Research*, **58**, pp. 129–52.
Wiseman, R. (1993a), 'Palladino and Barrington: Ten Major Errors', *Journal of the Society for Psychical Research*, **59**, pp. 16–34.
Wiseman, R. (1993b), 'Palladino and Fontana: Nine Major Errors', *Journal of the Society for Psychical Research*, **59**, pp. 35–47.

Wiseman, R. (1993c), 'Martínez-Taboas, Francia and Palladino: Nine Major Errors', *Journal of the Society for Psychical Research*, **59**, pp. 130–40.

Zusne, L. and Jones, W.H. (1982), *Anomalistic Psychology: A Study of Extraordinary Phenomena of Behavior and Experience*, Hillsdale, NJ: Lawrence Erlbaum.

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

Part I
Outside the Lab:
Paranormal Beliefs and Experiences

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[1]

A Community Mail Survey of Psychic Experiences

JOHN PALMER[1]

ABSTRACT: In March, 1974, a 46-item questionnaire was mailed to a randomly selected sample consisting of 300 students from the University of Virginia and 700 other adult residents of Charlottesville and surrounding suburbs. Items requested respondents to report the incidence and detailed characteristics of various psychic and psi-related experiences. Information concerning attitudes and the personal impact of such experiences was solicited, as well as demographic data.

Usable questionnaires were obtained from 89% of the student sample and 51% of the town sample. Claims of psychic and psi-related experiences were rather widespread: over half of the respondents claimed at least one ESP experience, for example. There also was a tendency for persons who claimed psychic or psi-related experiences to claim a large number of them. Variables related to naturally-occurring altered states (e.g., vividness and frequency of dream recall) tended to be strong predictors of such experiences, while demographic variables generally were poor predictors. However, there was a strong negative relationship between age and claims of *déjà vu* experiences. Many respondents indicated that psychic or psi-related experiences affected their attitudes toward life and/or life-styles.

INTRODUCTION

There have been a number of surveys of spontaneous psychic experiences reported in the parapsychological literature. However, most of these surveys involved preselected samples that might be atypical of a broadly representative population. L. E. Rhine (1961), for example, based her findings on the reports of persons who mailed descriptions of their experiences to her on their own initiative or in response to public appeals. Other studies involved asking questions about specific experiences to more or less intact groups

[1] This survey was conducted while I was Research Associate at the Division of Parapsychology, School of Medicine, University of Virginia. I wish to thank Dr. Ian Stevenson, Director of the Division, for providing financial and administrative support, and the Parapsychology Foundation for additional financial support.

such as college students or persons from a particular social class (Green, 1967; Prasad and Stevenson, 1968; Sidgwick and Committee, 1894; West, 1948). Perhaps the most representative study with an American sample is a national interview survey of mystical experiences which unfortunately dealt only superficially with psychic experiences (McCready and Greeley, 1976). A highly representative national survey of psychic experiences in Iceland, using a questionnaire similar to my own, has recently been published (Haraldsson, Gudmundsdottir, Ragnarsson, Loftsson, and Jonsson, 1977).

In 1974, I decided to undertake a survey of psychic experiences in the U.S. using random sampling techniques. My colleague in this endeavor was Mr. Michael Dennis.[2] A national survey was beyond our resources, so we decided on a community mail survey of Charlottesville, Virginia, and surrounding suburbs. Charlottesville is a community of about 35,000 people with a diversity of social and economic groups, although many of its resources are tied to the University of Virginia, which is located there. We nevertheless felt that Charlottesville is a reasonably representative American community, and a businessman of my acquaintance informed me that it is often considered such for purposes of marketing research.

Our objectives in carrying out the survey were to estimate the proportion of Americans who claim to have had various kinds of psychic expériences, and to explore correlations between these experiences and other variables, including related experiences and activities, attitudes, and demographic factors.

I want to stress at the outset that the survey dealt with experiences that our respondents *claimed* to have been psychic. I am not prepared to state what percentage of these cases actually require paranormal explanations, and to my knowledge no attempts have been made to verify any of them. I nevertheless think that the information obtained in the survey is of value to parapsychology both as a source of sociological information and of hypotheses about the nature of the experiences considered.

Method

Questionnaire

The questionnaire consisted of 46 items,[3] many of which contained several parts. Respondents answered by circling a number

[2] A preliminary report of this survey, co-authored by Mr. Dennis, was presented at the Seventeenth Annual Convention of the Parapsychological Association, Jamaica, N.Y. 1974. An abstract was published in *Research in Parapsychology 1974*. Metuchen, N.J.: Scarecrow Press, 1975. Pp. 130–133.

[3] For reasons of space, not all of these items will be discussed in this report.

next to their choices. They were encouraged to elaborate upon their answers or describe particularly meaningful experiences on the back of the questionnaire or on separate pages. The items can be classified in six main categories, as follows:

IA: Experiences that, if valid, by definition involve psi, i.e., either ESP or PK. These include waking ESP experiences, ESP dreams, being an "agent" for someone else's ESP experience, and poltergeist activity (RSPK).

IB: Experiences that are not psychic as such, but are of interest to parapsychologists because they might provide a context for either ESP or PK effects. These include out-of-body experiences (OBEs), apparitions, communication with the dead, hauntings, "memories" of a previous lifetime, *déjà vu* experiences, and aura vision. Supplementary questions (see below) explored possible psychic elements in some of these experiences.

II: Altered states of consciousness that are not of direct interest to parapsychologists, but are often considered relevant to psi. These include dreams (addressed in terms of frequency of recall and vividness), lucid dreams, and mystical experiences. The distinction between this category and category IB above is admittedly not a sharp one. For example, some might want to include mystical experiences in IB, while others might want to include *déjà vu* and aura vision in II. Our decisions on this matter were based on our assessment of which topics parapsychologists have historically considered to fall within their purview. They are to some extent arbitrary.

III: Activities related to psi. These include meditation, use of hallucinogenic drugs, analysis of one's dreams, and seeking the services of a psychic.

IV: Attitudes related to psi. Included in this category are attitudes toward astrology, survival of death, reincarnation, and the value of parapsychological research. We did not include a direct question about belief in psi (i.e., a "sheep-goat" question) because we feared it might bias respondents' answers to some of the other questions. We hoped that the question about the value of psi research would get at this issue indirectly, so we interpreted it as a surrogate sheep-goat question.

V: Demographic questions. These include sex, race, age, birth order, marital status, political ideology, religious denomination, religiosity, level of education, occupation, and family income.

VI: The effect of psychic experiences on the respondents' lives, including their attitudes, their life decisions, and whether such experiences had ever saved them or someone else from a crisis or tragedy.

For purposes of analysis, items in category I (A and B) were

treated as dependent or criterion variables, while items in categories II through V were treated as independent or predictor variables. Items in category VI have not been analyzed in relation to other items in the survey.

Items in categories I through III generally had multiple parts. First, respondents answered "yes" or "no" to whether they ever had the experience or engaged in the activity in question. If they answered "yes," they then answered a set of from one to eight supplementary questions asking for more specific information about these experiences. The first of these questions usually referred to how many times respondents had the experience, and they responded by circling a number from "1" to "9-or-more." For the other questions, they circled the number of these experiences that had a given characteristic. We recognized that in cases where persons had multiple experiences, they may well not be able to remember the exact number they had, or exactly how many had a given attribute. This format nonetheless allowed us to make rough estimates of these numbers.

The primary questions were phrased as precise descriptions of the experience or activity of interest, using the simplest words possible. In most cases, we avoided the use of labels such as "telepathy," "apparition," and "out-of-body experience" that might have different connotations for different respondents. We also generally avoided giving examples, lest respondents feel they should answer "no" unless their experience was the same as the example. These principles should become clear as I quote the questions in the presentation of results to follow.

Selection of Sample

Our purpose was to obtain as representative and random a sample as possible of the population of Charlottesville. Toward this end, we used two sources. The first was the City Directory of Charlottesville, which lists all persons over 18 living in numbered street addresses in Charlottesville and surrounding suburbs. The second was the University of Virginia (UVa) Student Directory, which gives a complete listing of students registered at the University at the beginning of the school year.

We decided upon an initial sample of 1,000 persons. Based upon census figures of the proportion of Charlottesville residents who were UVa students, we selected 700 names from the City Directory and 300 names from the Student Directory. These became our "town" (T) sample and our "student" (S) sample, respectively.

The names to be selected for the samples were determined by referring to a computer-generated table of pseudo-random num-

bers. These numbers defined the page, the column, and the row of the Directory which was to be sampled. This procedure became quite complicated with the City Directory, because of the nonsystematic way in which names were arranged on the pages. Also, names had to be excluded from each Directory for various reasons (e.g., UVa students had to be excluded if their names were "chosen" from the City Directory).[4]

Procedure

On March 1, 1974, one copy of the survey questionnaire, along with a postage-paid "business reply" return envelope, was sent to each of the 700 persons sampled from the City Directory. The first mailing to the 300 students was on March 11, because the week of March 1 coincided with the University's spring vacation. It also corresponded to a postal rate increase, which is why we didn't wait until March 11 to mail all the surveys.

There were two additional follow-up mailings to persons who had not yet returned their questionnaires. Each of these mailings occurred three weeks after the preceding one. It consisted of a new copy of the questionnaire, a new return envelope, and a supplementary letter exhorting the person to return his or her completed questionnaire.

Each questionnaire had a three-digit code number stamped on the lower right-hand corner of the back page. This number keyed the person's name on our mailing list. When a questionnaire was returned, this number was circled on the mailing list and the date we received it was recorded.

If a questionnaire was returned to us by the post office as undeliverable, or if someone else returned it indicating the person was deceased, a new name was sampled and the questionnaire was sent to the new individual. If a person returned an uncompleted questionnaire or indicated a refusal to cooperate, we simply treated that person as a "no-return" and did not resample.

Data Scoring and Analysis

The respondents' answers were transferred directly from the questionnaires to IBM cards by professional keypunchers at the UVa Computer Center. The data were subsequently stored on magnetic tape.

The data were analyzed using the SPSS statistical package (Nie,

[4] A manuscript describing the selection procedure in more detail is available from the author upon request.

Hull, Jenkins, Steinbrenner, and Bent, 1975). Frequency distributions were first printed out for all items. Most of the items were then cross-tabulated with each other, resulting in the printout of a large number of contingency tables and corrected chi-square values. In some cases it was necessary to combine some of the response categories for meaningful cross-tabulations.

RESULTS

Return Rates

We obtained usable questionnaires from 354 townspeople and 268 students, corresponding to 51% and 89% of the initial samples. Not all respondents answered every item, so that the "Ns" for individual items discussed below are often slightly less than the above figures.

We were very gratified by the response of the students, and I consider our final sample to be highly representative of this aspect of our population. The response of the townspeople, while less gratifying than the response of the students, was by no means a bad showing for this type of survey. Nevertheless, the representativeness of this sample is questionable.[5] Although we have not undertaken a formal comparison with census figures, it is obvious that there is an under-representation of the lower socio-economic classes in our final T sample. This is understandable, because such persons may have had difficulty in understanding the questions or been adverse to "paperwork" generally. The seriousness of this bias is mitigated somewhat by the fact (to be discussed later) that socio-economic variables were not strongly correlated with the frequency of reported psychic experiences.

A second way in which we attempted to assess the bias produced by the relatively low response rate of the townspeople was to evaluate the responses on questionnaire items as a function of when the respondents returned their questionnaires. Respondents were divided into three groups according to whether they returned their questionnaires after the first, second, or third mailings. For the T sample, these three groups contained 183, 112, and 59 respondents, respectively.

The original rationale for this procedure was based upon the assumption that persons who did not return their questionnaires at all were more like the persons who returned them after the third mailing than those who returned them after the first mailing. Thus if there were, say, a significant decline in the proportion of respon-

[5] I understand that sociologists consider 60% to be the minimal rate of return which justifies a claim of representativeness.

dents who reported having had a psychic dream across the three groups, we might suspect that the non-respondents had relatively few psychic dreams and that our sample percentage was an overestimate of the population percentage.

What we found, in fact, was that for none of the questionnaire items was there a significant difference in responses as a function of date of return. Moreover, the general trend was for a slightly higher proportion of people in the first and third groups to report psychic and psi-related experiences than in the second group. Although I cannot provide any evidence justifying the rationale outlined in the preceding paragraph, these results do make me more confident that the results from our T sample are not grossly off base.

Psychic Experiences

In this section I will present descriptive data regarding the experiences listed in Category I as defined above. These data are listed in Table 1. The figures in parentheses refer to the estimated proportion of *experiences* that have the characteristic in question.

A few comments about these latter estimates are in order. They were computed by dividing the total number of experiences reported as having the characteristic by the total number of experiences reported by respondents in the sample. Thus the experience rather than the respondent is the unit of analysis, and some respondents contributed to this figure more than others. Moreover, it was necessary to exclude from these computations the data from respondents who claimed to have nine or more of the experiences, since the exact number of experiences they had could not be determined. The adjacent figures not in parentheses refer to the proportion of those persons claiming to have had the experience at least once who also claimed to have had at least one such experience with the characteristic in question. The advantages and disadvantages of these figures are roughly complementary to those of the figures in parentheses.

For reasons of space, I will not quote all the figures in the text. Therefore, readers may wish to keep Table 1 at hand as they read the following paragraphs.

Waking ESP Experiences. This question was designed to assess how many respondents ever had what they considered to be a valid ESP experience while in the waking state. The question was worded as follows: "Have you ever had, *while awake*, a strong feeling, impression, or 'vision' that a previously unexpected event had happened, was happening, or was going to happen, and [learned] later that you were right?"

Table 1

PERCENTAGE OF RESPONDENTS CLAIMING PSI OR PSI-RELATED EXPERIENCES

Item	T (N=354)[a]		S (N=268)[a]	
Waking ESP	38		39	
More than one		86[b]		79[b]
Vision (hallucinations)		45 (30)[c]		24 (13)[c]
Tragic event		42 (21)		35 (17)
Family member		78 (51)		59 (33)
Within 24 hours		83 (56)		86 (67)
Told someone		55 (30)		46 (23)
ESP Dreams	36		38	
More than one		85		89
Especially vivid		81 (66)		80 (68)
Tragic event		43 (19)		31 (17)
Family member		65 (41)		58 (34)
Within 24 hours		63 (35)		58 (31)
Told someone		46 (22)		40 (16)
ESP Agency	18		20	
More than one		72		65
Emotion		62 (49)		58 (45)
Thinking of percipient		62 (49)		51 (43)
Family member		61 (47)		42 (32)
RSPK (Poltergeist)	8		6	
More than one		86		63
Other person present		46		27
Out-of-Body Experiences	14		25	
More than one		87		82
Saw physical body		56 (43)		62 (45)
Traveled		29 (21)		27 (14)
ESP (information acquired)		15 (7)		12 (3)
Seen as apparition		10 (9)		9 (2)
Produce at will		16 (12)		22 (16)
Apparitions	17		17	
More than one		74		79
Seen		46 (34)		41 (33)
Heard		70 (58)		65 (45)
Touched		61 (53)		56 (43)
Family member		49 (42)		29 (17)
Deceased		59 (39)		30 (19)
ESP (information acquired)		24 (18)		11 (2)
Collective		22 (12)		29 (13)
Communication with the Dead	8		5	
Seen or heard		72		31
Automatic writing		12		0
Direct voice		21		23
Xenoglossy		8		31
Lived in Haunted House	7		8	
Past-Life Memories	8		9	
More than one experience		69		87
Dream		65 (42)		68 (38)
More than one lifetime		36		32
Famous person		37		43

Table 1 (continued)

Item	T (N=354)[a]	S (N=268)[a]
Recalled details	32	27
Verified details	12	5
Déjà Vu	68	88
More than one	97	99
Aura Vision	5	6
More than one	87	67
See at will	37	47

[a] Sample size varies slightly from question to question due to non-responders.
[b] Figures in this column refer to the percentage of respondents who claimed the basic experience, not the percentage of the total sample.
[c] Figures in parentheses refer to the percentage of total experiences reported by respondents having the characteristic in question, excluding data from those reporting nine or more experiences.

At least one such experience was reported by 38% of the T sample and 39% of the S sample. The experiences tended to be intuitive rather than visual or hallucinatory in both samples, although a significantly greater percentage of these respondents reported at least one visual experience in the T sample than in the S sample (45% vs. 24%; $p < .01$).[6]

Contrary to what one might expect, only about one of five cases concerned "tragic events such as accidents or deaths," although about 40% of those reporting waking ESP experiences said they had at least one experience of this type. A large percentage of the cases were reported as involving members of the respondents' family, but these were a majority only in the T sample. The percentage of these respondents reporting at least one experience involving a family member was significantly greater in the T sample than in the S sample (78% vs. 59%; $p < .01$).

In a clear majority of the cases, the verifying event occurred within 24 hours before or after the experience, and in over a quarter of the cases the respondents said they told someone of their experience "before learning of the event by normal means." Thus a number of the cases are at least potentially verifiable.

ESP Dreams. This question was phrased as follows: "Have you ever had a rather clear and specific *dream* which matched in detail an event which occurred before, during, or after your dream, and which you did not know about or did not expect at the time of the dream?"

The figures for this item are very similar to those for the item concerning waking experiences. At least one psychic dream was

[6] Details of the chi-square analyses referred to here and elsewhere in this report are available from the author upon request.

reported by 36% of the T sample and 38% of the S sample. It is noteworthy that about two-thirds of these dreams were reported as being "more vivid" than the respondents' "ordinary dreams."

The figures regarding the subject matter of the dream (i.e., concerning a tragic event or a family member) are comparable to those reported for waking ESP experiences. However, the percentage of ESP dreams involving a family member was somewhat less than it was for waking experiences in the T sample (41% vs. 51%), and the percentage of townspeople reporting at least one such experience is not significantly different from the student percentage.

The data regarding potential veridicality (i.e., time interval between experience and verifying event, sharing the experience prior to verification) are comparable to the data for waking experiences.

ESP "agency." This question was designed to determine whether respondents believed they had ever been the object of someone else's ESP experience. Specifically, we asked: "Has any other person ever told you they had a dream, 'vision,' or definite feeling in which they seemed to get information about an event involving *you* which they could not have gotten in any 'normal' or conventional way?" Although we refer to this process as "agency," it should not necessarily be concluded that the "agent" played an active role in the experience.

Fewer of our respondents claimed to have been the source of an ESP "message" than claimed to have been on the receiving end. The above question was answered affirmatively by only 18% of the T sample and 20% of the S sample.

In a little less than half of the cases, it would appear that the respondent was "experiencing strong emotion at the time," and in about the same percentage he or she was "thinking of the other person at the time."

As was the case with receptive ESP experiences, a sizeable percentage of these "agency" cases involved family members, approaching half in the T sample. However, the difference between the T and S samples in the percentage of these respondents reporting at least one case of agency involving a family member did not reach significance.

RSPK. This question was designed to determine how many respondents were witnesses to poltergeist disturbances. The question was worded as follows: "Have you ever seen an object move with no 'natural' or physical means of motion that you could discover?"

Such an experience was claimed by 8% of the T sample and 6% of the S sample. It is difficult to determine how many of these persons could be identified as actual poltergeist agents. We attempted to get a handle on this by asking "whether these experiences occurred most often in the company of another person."

This seems to have been true in about a third of the cases overall, implying (albeit loosely) that the respondent was the agent in the other two-thirds.

Psi-Related Experiences

The data for psi-related experiences are also listed in Table 1.

Out-of-Body Experiences (OBEs). This question was worded as follows: "Have you ever had an experience in which you felt that 'you' were located 'outside of' or 'away from' your physical body; that is, the feeling that your consciousness, mind, or center of awareness was at a different place than your physical body? (If in doubt, please answer 'no.')"

At least one OBE was claimed by 14% of the T sample and 25% of the S sample. This difference is significant ($p < .01$). Our overall percentage, however, is similar to that obtained in other surveys (e.g., Green, 1967; Hart, 1954).

The respondents who reported OBEs said they saw their physical bodies while "outside" in about 44% of them. Thus while this is a common characteristic of OBEs, it is by no means universal. In only about one case out of six did respondents " 'travel' to distant places and 'see' or 'hear' what was going on there."

The percentage of OBEs for which the respondents claimed any evidentiality was small. I found it interesting that the number of claims of appearing as an apparition was almost identical to the number of claims of ESP associated with the experience. In no case is there firm evidence that these two attributes occurred together, but this is a possibility in six cases.[7] It is noteworthy that almost 20% of the respondents reporting OBEs said they had induced at least one voluntarily.

Apparitions. The question about apparitions was adopted essentially verbatim from the S.P.R. Census of Hallucinations (Sidgwick and Committee, 1894). It read, "Have you ever had, *while awake*, a vivid impression of seeing, hearing, or being touched by another being, which impression, as far as you could discover, was not due to any external physical or 'natural' cause?" (A sentence following this question cautioned respondents *not* to include experiences of the Christ or other religious figures.)

This question was answered affirmatively by 17% of the respondents in each sample, which is comparable to the uncorrected

[7] These were cases in which the respondents claimed to have had more than one OBE, and at least one with each of these attributes. Firm evidence could only be claimed if the respondent reported both attributes and only one OBE. There were no such cases.

figure of 13% obtained from the S.P.R. Census. Our results differed from the Census in one important respect, however. While the majority of the apparitions in the Census were visual, ours were predominantly auditory or tactile. Our results correspond more closely to those of L. E. Rhine (1956), who found that most of the "hallucinations" in her sample were auditory.

About 42% of the apparitions reported by the townspeople involved a family member, while only 17% of the students' apparitions fell into this category. However, the percentages of respondents in the two samples (49% vs. 29%) who saw at least one apparition of a family member did not differ significantly. The townspeople were about twice as likely as the students to "recognize the being as someone you knew had died sometime earlier" in at least one apparitional experience, and this time the difference is significant (59% vs. 30%; $p < .02$).

Only about 10% of the apparitions (mostly from the T sample) informed the respondent "of an unexpected accident or death." (But two of the five students in this group claimed to have had such an experience eight or more times!) Only about one in eight apparitions were collectively perceived, although approximately a quarter of those reporting apparitions claimed at least one collective experience. In the S.P.R. Census, where only one experience per person was analyzed, approximately a third of the apparitions were collectively perceived when more than one person was present.

Communication with the Dead. This question was designed to tap mediumistic-type experiences, although it might have overlapped somewhat with the question on apparitions. The question was worded as follows: "Have you ever 'communicated' with the dead or believed yourself to have been controlled or 'possessed' by a 'spirit'?"

This question was answered affirmatively by 8% of the T sample and 5% of the S sample.

Among the 26 townspeople who responded affirmatively, seeing the "spirit" or hearing its voice was the most common mode of communication (18 persons), although three reported control of handwriting (automatic writing), five reported that a "spirit" had controlled their voices (direct voice), and two reported that it had communicated through them in a foreign language they did not know (xenoglossy).

Among the 13 students who claimed communication with the dead, four said they had seen or heard a "spirit'," three reported direct voice, and four reported xenoglossy. None reported automatic writing.

Haunting. This question asked respondents, "Have you ever lived in a house you believed to be 'haunted'?" An affirmative

response was given by 7% of the T sample and 8% of the S sample.

Past-Life "Memories." We asked our respondents: "Have you ever had what seems to be a 'memory' of a previous lifetime (i.e., reincarnation)?" Such "memories" were claimed by 8% of the T sample and 9% of the S sample.

Approximately two-thirds of those who recalled previous lifetimes said that at least one of their memories came in the form of a dream, although only 40% of the total memories reported were dreams.

About a third of our recallers claimed to remember more than one lifetime, while about 40% remembered being a "well-known or important person."

Fourteen respondents (30% of all recallers) said they were "able to recall details such as names, places, historical events, etc.," which they hadn't known about prior to experiencing the "memories," but only four of them claimed to have "checked any such details and found them to be correct in records of the actual past."

Déjà Vu. This question was phrased as follows: "Have you ever had the strong feeling or impression that you had been some place or in the same situation before, even though you had never actually been there before or were experiencing the event for the first time in 'real life'?"

This question evoked a far higher percentage of affirmative responses than did any of the other questions relating to the supposedly psi-related experiences we examined. It was answered affirmatively by 68% of the T sample and 88% of the S sample. This difference is highly significant ($p < .001$).

Aura Vision. We asked, "Have you ever seen light or lights around or about a person's head, shoulders, hands, or body which, as far as you could tell, were not due to 'normal' or 'natural' causes (i.e., 'halo' or 'aura')?"

This question was answered affirmatively by 5% of the T sample and 6% of the S sample. Over 40% of these positive responders said they could see the light around people when they wanted to, at least sometimes.

Multiplicity of Psi or Psi-related Experiences. There is considerable evidence from the survey that most of the respondents who claimed to have had psychic or psi-related experiences apparently had a large number of them. First, most respondents who reported one experience of a given type said they had more than one of this type. These percentages range from a low of 63% to a high of 99% (see Table 1). Reports of having had such experiences nine or more times were quite common.

Secondly, there was a tendency for respondents who reported

one type of experience to report other types as well. This was especially true in the T sample. Among the 55 cross-tabulations contrasting the 11 items discussed so far two at a time, 50 of them were statistically significant (i.e., $p < .05$), most highly so. This consistency was somewhat less evident in the S sample, where only 26 of the 55 cross-tabulations were significant, but it is still present to a high degree. The discrepancy between the samples is not primarily attributable to the smaller number of respondents in the S sample. In 43 of the 55 tabulations, the magnitude of the percentage differences within the chi-square table was greater in the T sample than in the S sample.

In an analysis in which mystical experiences were included among psi-related experiences, we found that 82% of the T sample and 96% of the S sample had at least one psychic or psi-related experience. Fifty-eight percent of the T sample and 74% of the S sample reported more than one. Seventeen percent of the T sample and 19% of the S sample reported five or more.

Altered States of Consciousness

Frequency data for items in this category are presented in Table 2 and their relationships with psi-related experiences in Figure 1. All relationships discussed in this and the next two sections are in the positive direction.

Table 2

PERCENTAGE OF RESPONDENTS HAVING EXPERIENCES OR PRACTICING ACTIVITIES OF RELEVANCE TO PSI

Item	T (N=354)[a]		S (N=268)[a]	
Recall of Dreams				
Once a week or more	42		67	
Once or twice a month	24		19	
Rarely or never	34		14	
Vivid Dreams				
Once a week or more	29		44	
Once or twice a month	30		33	
Rarely or never	41		24	
Lucid Dreams				
Once a week or more	4		8	
Once or twice a month	10		21	
Rarely	42		42	
Never	44		29	
Mystical Experiences	28		35	
More than one		79[b]		69[b]
Self-Analysis of Dreams	36		53	
How long				
Less than 6 months		32		34

Table 2 (continued)

Item	T (N=354)[a]	S (N=268)[a]
6 months to 5 years	22	48
More than 5 years	45	19
How helpful		
Very helpful	13	2
Somewhat helpful	46	47
Not helpful	37	50
Harmful	3	1
Sought a Psychic	10	3
More than once	72	38
More than one psychic	45	0
Acted upon advice	33	78
How helpful		
Very helpful	15	67
Somewhat helpful	30	22
Not helpful	52	11
Harmful	3	0
Meditation	6	9
How long		
Less than 6 months	42	44
6 months to 5 years	21	56
More than 5 years	37	0
How helpful		
Very helpful	44	23
Somewhat helpful	50	59
Not helpful	0	18
Harmful	6	0
Mind-Expanding Drugs	7	32
How frequently		
Once a week or more	33	37
Occasionally	17	20
Seldom	28	22
Once or twice	22	22
Psi-related experiences during	29	28

Note; figures in some percentage columns do not add up to 100 due to rounding error.

[a] Sample size varies slightly from question to question due to non-responders.

[b] Figures in this column refer to the percentage of respondents who claimed the basic experience or activity, not the percentage of the total sample.

Dreams. Three of the four questions in the "altered states" category concerned dreams. Forty-two percent of the T sample and 67% of the S sample stated that they recalled their dreams at least once a week; 29% of the T sample and 44% of the S sample stated that at least once a week they had a dream they would describe as "vivid."

Lucid dreams, which we defined in the questionnaire as "a special sort of dream in which you knew *during the dream* that you

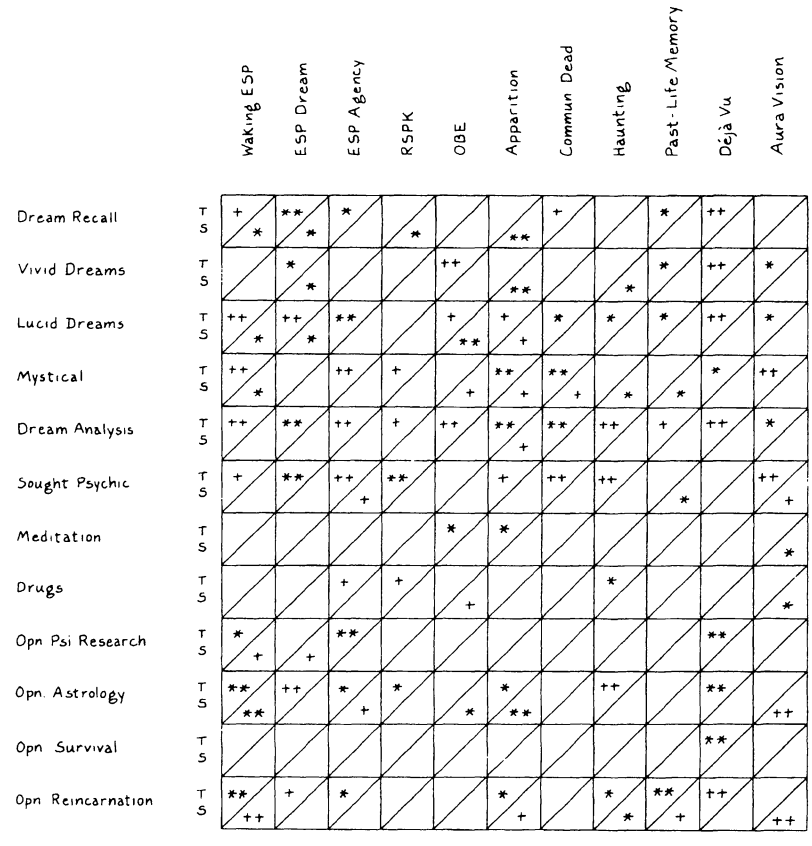

Fig. 1. Relationships between psi-related experiences and other experiences, activities, and opinions (*$p < .05$; +$p < .01$; **$p < .001$; ++$p < .0001$).

were dreaming and felt that you possessed all your waking faculties," were claimed by 56% of the T sample and 71% of the S sample. However, only 14% of the T sample and 29% of the S sample said they had such dreams more often than "rarely."

These items were generally good predictors of psi-related experiences,[8] especially ESP experiences. Frequency of dream recall was significantly related to waking ESP experiences and ESP dreams in both samples. In the T sample, it was also related significantly to

[8] From here on, the term "psi-related experiences" will also include experiences distinguished as "psychic"; i.e., those in category IA as well as IB.

ESP agency, communication with the dead, past-life memories, and *déjà vu*. On the other hand, it related significantly only to RSPK and apparitions in the S sample.

Vividness of dreams was not quite as reliable a predictor as frequency of dream recall. The only variable it related to significantly in both samples was ESP dreams, a finding that should be interpreted in the context of the fact reported above that respondents tended to rate their ESP dreams as more vivid than their ordinary dreams. Vividness also was significantly related to OBEs, past-life memories, *déjà vu,* and aura vision in the T sample, and to apparitions and hauntings in the S sample.

Whether or not a respondent had ever had a lucid dream was a strong predictor of psi-related experiences. This was especially true in the T sample, where it was significantly related to every psi-related experience except RSPK. In the S sample, it was significantly related only to waking ESP, ESP dreams, OBEs, and apparitions.

Mystical Experiences. The survey revealed that 28% of the T sample and 35% of the S sample claimed to have had "a profound and deeply moving 'spiritual,' 'mystical,' or transcendental experience." Of those claiming a mystical experience, about three-quarters said they had had more than one.

Mystical experiences were significatly related to waking ESP in both samples, but to ESP dreams in neither. This item also significantly predicted apparitions and communication with the dead in both samples. In addition, it significantly predicted ESP agency, RSPK, *déjà vu,* and aura vision in the T sample, and OBEs, hauntings, and past-life memories in the S sample.

Activities Related to Psi

Data for items in this section are also in Table 2 and Figure 1.

Dream Analysis. We asked our respondents the following: Have you "ever tried to remember or analyze your dreams for the guidance or insight they might give you?" This question was answered affirmatively by 36% of the T sample and 53% of the S sample. Fifty-nine percent of these respondents in the T sample found such analysis to be at least somewhat helpful, as compared to 49% of the S sample.

I was interested in this as a predictor of psi experiences, especially because it conceivably might separate out persons who had been in psychotherapy, and whose experiences might have some pathologic origin. We did not wish to approach this question directly for obvious reasons. Of course, an affirmative answer to this

question by no means necessarily implies psychotherapeutic experience.

This item did prove to be a very strong predictor of psi experiences, but only for the T sample. There was a significant relationship between reported self-analysis of dreams and every one of the 11 psi-related experiences for the T sample. For the S sample, this relationship was only significant for apparitions, although the other relationships tended to be in the positive direction.

Visits to Psychics. We asked our respondents whether they had "ever seriously sought information, help, or guidance" from a "medium, clairvoyant, or psychic," "palm reader," "astrologer," or "faith (or psychic) healer." The question was answered affirmatively by 10% of the T sample and 3% of the S sample. Interestingly, while 89% of those in the S sample who sought such help found it at least somewhat helpful, only 45% of the corresponding respondents in the T sample did.

Again, this item proved to be a better predictor of psi-related experiences in the T sample than in the S sample. It was significantly related in this sample to every psi-related experience except OBEs, past-life memories, and *déjà vu*. In the S sample, it was significantly related only to ESP agency, past-life memories, and aura vision. The extreme marginal totals in these latter analyses, however, suggest some caution in their interpretation.

Meditation. We asked our respondents whether they had ever practiced meditation, in the sense of a "*formal* technique of stilling the mind." The question was answered affirmatively by only 6% of the T sample and 9% of the S sample. Meditation was described as at least somewhat helpful by 94% of those in the T sample who practiced it, and 82% of those in the S sample.

Meditation was not a strong predictor of psi-related experiences, although the direction of the relationship was almost always positive. It was significantly related to OBEs and apparitions in the T sample and to aura vision in the S sample.

Drugs. We asked our respondents whether they had "ever used 'mind expanding' drugs or medicines," and, if so, whether they had had any psi-related experiences while under their influence. We used this rather non-specific wording to avoid putting some of our respondents in the position of admitting the use of illegal substances. Although our survey was essentially anonymous, we did keep records of the names associated with the code numbers on the questionnaires until the solicitation phase of the project was completed.

The drug question was answered affirmatively by 7% of the T sample and 32% of the S sample. We suspect that the figure for the S sample is a gross underestimate of actual drug use. Whether this

resulted from a fear of legal difficulties, whether the term "mind expanding drugs" was interpreted by some as not including the milder hallucinogens such as marijuana, or whether some other factor is involved is unclear.

In any event, drug use was a surprisingly poor predictor of psi-related experiences, although relationships were generally positive. In the T sample, it was significantly related only to ESP agency and to witnessing of RSPK and haunting phenomena. (Again, caution is required here in interpreting this finding because of the extreme marginal totals on both variables.) In the S sample, drug use was significantly related to OBEs and aura vision. We suspect that the drug-OBE relationship is the explanation of why OBEs were significantly more prevalent in the S sample than in the T sample.

It also should be mentioned, however, that these relationships do not directly measure whether respondents who used drugs had psi-related experiences *while under the influence* of such drugs. Only 29% of the drug users said they had at least one psi-related experience under the influence of drugs; and 13% of the T sample who reported OBEs and 21% of the S sample who reported OBEs said they had at least one OBE under the influence of drugs.

Opinions on Psi-Related Topics

Frequency data for the opinion items are presented in Table 3 and their relationships to psi-related experiences in Figure 1.

Parapsychological Research. We asked our respondents to choose among a series of alternatives concerning their opinions about "the scientific study of psychic phenomena." The responses indicated that they had an overall positive attitude toward parapsychological research. Most respondents opted for the moderately positive response that since "mankind has recorded a long history of psychic experiences," science should study them so that "we may be able to discover some basic facts about them." However, only 11% of the T sample and 17% of the S sample agreed that "the discovery of the laws governing psychic phenomena will be one of the most important discoveries in the history of science." On the other hand, only 8% of the T sample and 9% of the S sample agreed that if psychic phenomena exist they are "probably of little value to mankind because they are so unpredictible" and that "science can probably not either prove or disprove them." Only a handful of respondents opted for the more extremely negative stance that the research is "a bunch of foolishness and nonsense" or that psychic phenomena "are probably the work of the devil."

As a predictor of psi-related experiences, this question related

Table 3

PERCENTAGE OF RESPONDENTS HOLDING VARIOUS OPINIONS ON PSI-RELATED TOPICS

Topic	T (N=354)[a]	S (N=268)[a]
Parapsychological Research		
Sinful	1	2
Foolish	3	0
Unfruitful	8	9
Meaningful	58	69
Important	11	17
No opinion	20	5
Astrology		
A certainty	3	2
A probability	8	6
A possibility	41	35
Unlikely	20	32
Untrue	17	24
No opinion	12	2
Survival		
A certainty	35	32
A probability	13	22
A possibility	29	27
Unlikely	8	13
Untrue	8	5
No opinion	7	3
Reincarnation		
A certainty	2	3
A probability	8	9
A possibility	29	34
Unlikely	24	31
Untrue	28	20
No opinion	10	3

Note: Figures in some percentage columns do not add up to 100 due to rounding error.

[a] Sample size varies slightly from question to question due to non-responders.

primarily to items dealing directly with ESP. It was significantly related to waking ESP, ESP agency and *déjà vu* in the T sample, and to waking ESP and ESP dreams in the S sample.

Astrology. Only 11% of the T sample and 8% of the S sample considered astrology to be probably or certainly true, while 37% of the T sample and 56% of the S sample considered it to be probably or certainly untrue.

Attitudes toward astrology did prove to be a relatively good predictor of psi-related experiences in both samples. It was significantly related to waking ESP, ESP agency, and apparitions in both samples. In addition, it was significantly related to ESP dreams, RSPK, haunting, and *déjà vu* in the T sample, and to OBEs and

aura vision in the S sample. In neither sample was it significantly related to communication with the dead or past-life memories.

Survival. Forty-eight percent of the T sample and 54% of the S sample considered survival of death to be either probable or certain. In contrast only 16% of the T sample and 18% of the S sample considered survival to be unlikely or not true.

Belief in survival proved to be a very poor predictor of psi-related experiences. The only significant relationship was with *déjà vu* in the T sample.

Reincarnation. The concept of reincarnation was considered to be probably or certainly true by 10% of the T sample and 12% of the S sample. It was considered unlikely or untrue by 52% of the T sample and 51% of the S sample.

Belief in reincarnation was a much better predictor of psi-related experiences than was belief in survival generally. The relationships with waking ESP, apparitions, hauntings, and (of course) past-life memories were significant for both samples. In addition, relationships with ESP dreams and ESP agency were significant for the T sample, and aura vision for the S sample.

Demographic Variables

Frequency data for these variables are listed in Table 4 and their relationships to psi-related experiences shown in Figure 2. Because these significant relationships are less frequent and also less obvious than those reported previously, they will be presented in somewhat greater detail.

Sex. The T sample was predominantly female (60%), while the S sample was predominantly male (66%). However, in neither sample was there a clear-cut sex difference for any of the psi-related variables. The closest approximation was the tendency for more females to have reported a waking ESP experience than males in the S sample (48% vs. 34%; $p = .052$).

Race. The T sample contained only 7% blacks and the S sample no blacks. There were too few Orientals in either sample to allow inclusion of this group in the analyses.

Generally speaking, whites and blacks did not differ in the proportions reporting psi-related experiences. However, a significantly higher percentage of blacks than whites reported having at least one ESP dream in the T sample (56% vs. 34%; $p < .05$).

Age. The age of our respondents in the T sample also proved to be a poor predictor of psi-related experiences. There was, however, a strong linear tendency for a greater percentage of younger persons than older persons to report *déjà vu* experiences ($p < .00001$). The range was from 83% for those under 30 to 52% for

Table 4

PERCENTAGE OF RESPONDENTS IN VARIOUS DEMOGRAPHIC CATEGORIES

Item	T (N=354)[a]	S (N=268)[a]
Sex		
Male	40	60
Female	60	34
Race		
White	92	98
Black	7	0
Oriental	1	2
Age		
30 and under	26	99
31–40	23	1
41–50	23	0
51–60	14	0
61–70	10	0
Over 70	3	0
Birth Order		
Only child	11	6
First born	27	27
Second born	23	30
Later born	39	37
Marital Status		
Single	9	83
Married	77	17
Separated or divorced	7	0
Widowed	7	0
Politics		
Very conservative	5	1
Conservative	22	11
Moderate	52	46
Liberal	18	34
Very liberal	3	8
Religion		
Protestant	74	50
Catholic	5	16
Jewish	2	3
Atheist	8	21
Other	11	10
Religiosity		
Very religious	8	9
Moderately religious	61	41
Slightly religious	20	28
Not at all religious	11	21
Education (Highest Level)		
Grade school	15	0
High school or trade school	28	0
Some college (undergraduate)	16	63
College graduate (Bach.)	22	0
Post-graduate	19	37
Occupation		
Student	4	100
Blue collar	11	0

Table 4 (continued)

Item	T (N=354)[a]	S (N=268)[a]
Clerical	16	0
Business	8	0
Professional (Bach. degree)	14	0
Professional (Advanced degree)	12	0
Housewife	23	0
Retired	7	0
Other	5	0
Income[b]		
Under $5,000	11	—
$5,000–$10,000	17	—
$11,000–$15,000	27	—
$16,000–$20,000	17	—
$21,000–$30,000	17	—
Over $30,000	11	—

Note: Figures in some percentage columns do not add up to 100 due to rounding error.

[a] Sample size varies slightly from question to question due to non-responders.

[b] Figures for students are not included because some respondents interpreted the item as requesting parental income, while others did not.

those over 50. A similar trend was reported by McCready and Greeley (1976). This pattern is reinforced in our data by the fact that the incidence of *déjà vu* in the S sample, most of whom were under 30, was 88% (similar to the percentage of the "under 30's" in the T sample).

Birth Order. To assess birth order, we classified respondents as only children, first born, second born, or later born. Generally speaking, birth order was not a significant predictor of psi-related experiences. The one exception was ESP agency in the S sample ($p < .05$). This difference is attributable to the fact that 50% of the 14 "only children" reported ESP agency as compared to only 18% of the others. However, the small number of only children, the marginal significance of the relationship, and the failure of this variable to be a significant predictor of other related variables, suggest caution in interpreting the finding.

Marital Status. Respondents were classified as single, married, separated/divorced, or widowed. In the T sample, there was a general tendency for either the separated/divorced respondents and/or the widowed respondents to report a greater number of psi-related experiences than single or married respondents, but this trend was only significant for five of the 11 experiences. These were waking ESP ($p < .05$), ESP dreams ($p < .05$), OBEs ($p < .05$), apparitions ($p < .05$), and hauntings ($p < .05$).

The widowed were responsible for the effect primarily for the

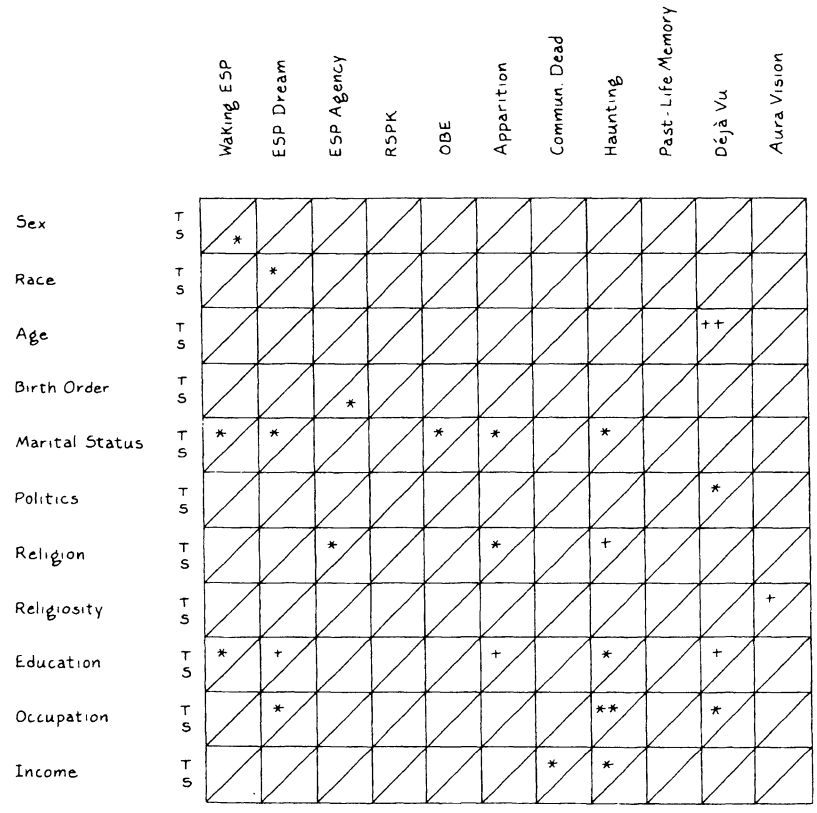

Fig. 2. Relationships between psi-related experiences and demographic variables (*$p < .05$; +$p < .01$; **$p < .001$; ++$p < .0001$).

survival-related experiences. Compared to the single and married (combined) groups, the widowed had a high incidence of ESP dreams (60% vs. 33%), apparitions (37% vs. 15%), and hauntings (22% vs. 6%). The separated/divorced group had a relatively high incidence of waking ESP (67% vs. 36%), ESP dreams (50% vs. 33%), and OBEs (35% vs. 13%).

Politics. Respondents were asked to rate themselves on a five-point scale from "very conservative" to "very liberal." There was a slight trend for conservatives to report more psi-related experiences than liberals, but this was somewhat variable and never significant. At best, it is a weak trend. The only significant relationship was a tendency for liberals to report more *déjà vu* experiences than conservatives in the T sample, ($p < .02$). Eighty-three

percent of those who rated themselves liberal or very liberal reported *déjà vu,* while only 61% of those who rated themselves conservative reported such experiences.

Religion. Respondents were classified as Protestant, Catholic, Jewish, atheist, or "other." In the T sample, the general trend was for those classified as "other" to report a relatively high incidence of psi-related experiences and those claiming to be atheists to report a relatively low incidence. Those classified as Jewish reported a low incidence in both samples. However, their numbers were only six and eight in the T and S samples, respectively, requiring extreme caution in drawing conclusions. These patterns were significant only for ESP agency ($p < .05$), apparitions ($p < .05$), and hauntings ($p < .01$).

Religiosity. Respondents rated themselves on a four-point scale from "very religious" to "not at all religious." Sixty-nine percent of the T sample and 50% of the S sample rated themselves as at least moderately religious. However, religiosity was a poor predictor of psi-related experiences, with no consistent directional trend evident. The only significant finding was a positive relationship between religiosity and aura vision in the T sample ($p < .01$).

The next three items all measure aspects of socio-economic status, and are highly confounded.

Education. The relationship between level of education and psi-related experiences in the T sample differed sharply for different types of experiences. Significant relationships were found with waking ESP ($p < .05$), ESP dreams ($p < .01$), apparitions ($p < .01$), hauntings ($p < .05$), and *déjà vu* ($p < .01$). For all these relationships except *déjà vu*, the higher incidence occurred among those with the least education.

Those who reported having no more than a high school education claimed a relatively high incidence of waking ESP and of apparitions. For waking ESP, the incidence in this group was 47% compared to 33% for the other respondents. The comparable figures for apparitions were 23% and 11%, this time including trade school graduates in the less educated group.

The effects for ESP dreams and hauntings were due primarily to those with just a grade school education. Sixty-one percent of this group reported at least one ESP dream, compared to 32% of the other respondents. The comparable figures for hauntings were 18% and 5%.

Déjà vu provided a reversal of this pattern, with the highest reported incidence (81%) among those with a graduate degree, and the lowest (48%) among those with only a grade school education.

There were no significant differences on any of the items between undergraduate and graduate students in the S sample.

Occupation. In the T sample, there were significant occupational differences for ESP dreams ($p = .05$), hauntings ($p < .001$), and *déjà vu* ($p < .05$). For ESP dreams and hauntings, there were noticeably high percentages among blue collar workers (50% and 23%, respectively). There was a low percentage of ESP dreams among respondents with advanced degrees (19%). For *déjà vu*, this pattern was reversed. There was a high incidence of this experience among professional persons with bachelor or advanced degrees (80%) and a relatively low incidence among blue collar workers (50%).

Income. Total family income before taxes in the T sample was significantly related only to communication with the dead ($p = .02$) and hauntings ($p < .02$). In both cases, the trend of the relationship was generally negative, with the highest incidence (15% and 18%, respectively) among those earning less than $5,000 per year. In both cases, however, there was a relatively high incidence (17% and 13%) among respondents earning between $16,000 and $20,000 per year that broke the general linear trend. Data on this variable are suspect in the S sample because it was not always clear whether respondents considered parents' income in their responses.

Sociological Factors

We asked our respondents whether any of their psychic experiences had ever saved them from "a serious or tragic event such as illness, severe emotional crisis, accident, or death." We also asked whether such an experience had ever saved another person in a crisis situation, or whether another person's psychic experience had saved the respondent. As can be seen in Table 5, less than 10% of the respondents in each sample answered these questions affirmatively. Nevertheless, the survey shows that a few respondents considered that at least one psychic experience with which they had been associated was useful in a practical way in a crisis situation.

Further, we asked our respondents whether any of the experiences they had reported had significantly influenced or changed any of their "feelings or attitudes," and also whether any of these experiences had significantly influenced or changed any of the "important decisions" they had made in their lives. As indicated in Table 5, the alternatives we listed for both these questions evoked varying responses. Among attitudes, those regarding the self, humanity, spiritual beliefs, life, and play were most frequently cited as having been changed. As for decisions, our data seem to indicate that psi experiences are most likely to affect one's life style, which

Table 5

PERCENTAGE OF RESPONDENTS REPORTING SOCIOLOGICAL SIGNIFICANCE FOR PSI EXPERIENCES

Item	T (N=354)[a]	S (N=268)[a]
Saved in Crisis		
Saved self	9	7
Saved someone else	9	5
Saved by someone else	7	4
Changed Attitudes		
Self	27	36
Humanity	17	26
Society	6	8
Spiritual beliefs	23	25
Nature	12	15
Life (meaning of)	25	31
Death	13	20
War	5	8
Sex, love	16	18
Family	17	16
Education	6	8
Business	6	7
Science	5	8
Wealth	9	14
Fame	5	13
Media	5	4
Art, leisure, play	15	25
Other	4	8
Influenced Decisions		
Friends	14	15
School	7	13
Military	2	2
Job	12	5
Vacation	6	6
Moving	4	3
Politics	3	1
Religious denomination	8	12
Marriage	8	6
Children (whether to have)	3	1
Naming child	2	1
Home	4	2
Car or appliance	2	2
Health	10	8
Life style	21	26
Other	2	4

[a] Sample size varies slightly from question to question due to non-responders.

is perhaps the most fundamental of the alternatives listed. At least one respondent checked every one of the alternatives for both questions.

Discussion

The most elementary, and perhaps the most important, conclusion that can be drawn from the survey is that a substantial percentage of the respondents considered themselves to have had interactions with their environment that cannot be accounted for by known physical laws. More than a third of our sample, for instance, responded affirmatively to the items concerning waking ESP experiences and ESP dreams. These figures are comparable to those reported by McCready and Greeley (1976) based upon similar questions. We calculated that 51% of our T sample and 55% of our S sample claimed either a waking ESP experience or an ESP dream.

We were also impressed by the large number of respondents who claimed multiple psi or psi-related experiences. It does not appear from our data that such experiences are normally distributed in the population. On the contrary, it seems that our population consisted of two rather distinct subgroups: those who claimed no or very few psychic experiences, and those who claimed a large number.

For the most part, the incidence of the various kinds of psi and psi-related experiences was quite similar in the T and S samples. The two exceptions were OBEs, probably attributable to greater drug use on the part of the students, and *déjà vu*, most likely attributable to the age factor. There was also a tendency for more of the experiences from the T sample to involve a family member or someone who had died (especially in the case of apparitions). The similarity of results from the two samples lends support to the validity of our findings from the T sample, whose return rate was lower than that of the S sample. This latter sample was quite representative of the UVa student population.

The other items in the survey varied widely in their utility as predictors of psi and psi-related experiences. Among the most successful predictors in both samples were the items concerning dreaming, especially frequency of dream recall and lucid dreaming. Both of these items reflect the degree to which the conscious mind is capable of gaining access to the content of the unconscious mind. The only difference is that in lucid dreaming the access is contemporaneous, while with dream recall it is retrospective. The apparent fact that this faculty makes one prone to psi or psi-related experiences supports the view that such experiences are latent in the unconscious mind and potentially available to those who can break through the repressive mechanisms (whatever their basis) that ordinarily prevent them from reaching consciousness. Of course, evidence derived from more veridical instances of psi is still needed.

Mystical experiences, which may well reflect the access to the unconscious in the waking state, was a significant predictor of waking ESP experiences (and other waking psi-related experiences), but not of ESP dreams.

Two activities that reflect a quest for deeper self-insight were significantly related to psi and psi-related experiences only in the T sample. These were analysis of one's own dreams and visits to psychic counselors. It is likely that psychic experiences prompt one to engage in such activities, and this seems to be the most plausible explanation of the relationship. Why these items were not better predictors in the S sample is unclear.

Generally speaking, drug use and meditation were poor predictors of psi and psi-related experiences in both samples. These results pertaining to induced altered states are in sharp contrast to the more positive results with natural altered states such as dreams and mystical experiences. There is little evidence from the survey that inducing altered states "opens one up" to psi or psi-related experiences. One is either predisposed to such experiences naturally, or one is unlikely to have them.

Attitudes concerning the importance of parapsychological research were good predictors of psi experiences but not of psi-*related* experiences. This is not surprising, since parapsychological research is only directly concerned with the former.

Opinions about survival of death, as well as the demographic item concerning religiosity, were unrelated to psi or psi-related experiences. This result also is not surprising since attitudes on these topics are conditioned by cultural factors having little to do with the paranormal. Likewise religious affiliation was not generally a strong predictor of psi or psi-related experiences, although there was a slight tendency for there to be a relatively low incidence of such experiences among atheists and Jews and a relatively high incidence among those who described their religious affiliation as "other." If this trend has any validity, it would make some sense. While differing from atheists in some respects, Jews tend to share with them a relatively skeptical attitude toward the more "esoteric" aspects of religion, while the "other" religions claimed by some respondents are most likely faiths (e.g., "Eastern" faiths) where such aspects are stressed.

Belief in reincarnation and in astrology, which in contrast to belief in survival are not culturally sanctioned and more reflective of occult predispositions, were relatively strong predictors of psi and psi-related experiences.

Demographic variables did not fare well as predictors in the survey, especially those that are biologically based. Sex, race, and birth order were related marginally only to waking ESP experi-

ences, ESP dreams, and ESP agency, respectively. I would put little weight on any of these relationships unless they are replicated. The tendency for *déjà vu* to be more prevalent among younger persons is a much more solid finding, but I can offer no satisfying explanation for it.

Whether respondents in the T sample were married or single had no effect on the incidence of psi or psi-related experiences. However, one of the more intriguing (although marginal) findings of the survey was the relatively high incidence of ESP and OBEs among respondents who were separated or divorced. One might speculate that the tension surrounding marital difficulties facilitates psi and related experiences, or, on the other hand, that psychic experiences and associated attitudes or life styles make people difficult to live with in a marital relationship. Obviously, more data are needed. The relative frequency of experiences possibly involving communication with the dead among widowed respondents is much less difficult to account for.

Socio-economic factors (education, occupation, family income) were not strongly associated with psi or psi-related experiences. However, there was a trend in the case of some items (ESP, apparitions, and hauntings) toward a greater incidence of such experiences among the lower socio-economic levels. It is unclear whether the relationship means that such persons actually have more experiences or whether they adopt a less critical attitude in interpreting them. Since the lower socio-economic levels were under-represented in our survey, this finding may also indicate that our estimates of the total proportion of persons claiming these experiences in our population are slightly conservative.

Ironically, *déjà vu* experiences were reported more frequently by the higher socio-economic levels in the T sample. *Déjà vu* also was the only item significantly related to political preference. Liberals were more likely to claim *déjà vu* experiences than conservatives, which is also opposite to the general trend for other experiences.

Finally, it is clear from the sociological questions that a substantial number of our respondents were affected in some way by their experiences. Whether or not such experiences are valid in the scientific sense, it is evident that they have impact on people's lives. If nothing else justifies greater efforts toward a scientific understanding of psychic phenomena, this fact certainly does.

References

GREEN, C. E. Ecsomatic experiences and related phenomena. *Journal of the Society for Psychical Research*, 1967, 44, 111–131.

HARALDSSON, E., GUDMUNDSDOTTIR, A., RAGNARSSON, A., LOFTSSON, J., AND JONSSON, S. National survey of psychical experiences and attitudes toward the paranormal in Iceland. In J. D. Morris, W. G. Roll, and R. L. Morris, (Eds.), *Research in Parapsychology 1976.* Metuchen, N.J.: Scarecrow Press, 1977. Pp. 182–186.

HART, H. ESP projection: Spontaneous cases and the experimental method. *Journal of the American Society for Psychical Research,* 1954, 48, 121–146.

MCCREADY, W. C., AND GREELEY, A. M. *The Ultimate Values of the American Population.* Beverly Hills, Calif.: Gage, 1976.

NIE, N. H., HULL, C. H., JENKINS, J. G. STEINBRENNER, K., AND BENT, D. H. *SPSS: Statistical Package for the Social Sciences.* (2nd ed.) New York: McGraw-Hill, 1975.

PRASAD, J., AND STEVENSON, I. A survey of spontaneous psychical experiences in school children of Uttar Pradesh, India. *International Journal of Parapsychology,* 1968, 10, 241–261.

RHINE, L. E. Hallucinatory psi experiences. I. An introductory survey. *Journal of Parapsychology,* 1956, 20, 233–256.

RHINE, L. E. *Hidden Channels of the Mind.* New York: Sloane, 1961.

SIDGWICK, H., AND COMMITTEE. Report on the census of hallucinations. *Proceedings of the Society for Psychical Research,* 1894, 10, 25–422.

WEST, D. J. A mass observation questionnaire on hallucinations. *Journal of the Society for Psychical Research,* 1948, 34, 187–196.

John F. Kennedy University
12 Altarinda Road
Orinda, California 94563

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[2]

Psychology and Coincidences

Caroline A. Watt
Department of Psychology, University of Edinburgh

Abstract: The paper presents a selective review of research suggesting possible normal causes for some coincidences. After a brief discussion of hidden causes, predictions with multiple endpoints, and simple probability, the bulk of the paper focuses on psychological research into judgement and decision-making under uncertainty. Shortcuts in information processing that have been held responsible for apparent weaknesses in everyday statistical intuitions are discussed, as are recent criticisms of this heuristics and biases paradigm. Examples are given of studies demonstrating how perception, judgement and recollection may be biased so as to confirm our preconceptions. Some implications of this research for the study of coincidences are pointed out, and research suggesting promising remedial measures to improve judgement is noted.

The *Little Oxford Dictionary* (1986) defines a coincidence as a 'remarkable concurrence of events without apparent causal connection'. This definition begs 2 questions: what makes some concurrences of events remarkable and not others, and how does one establish an apparent lack of causal connection? By their nature, remarkable coincidences are one-off, unique events that cannot realistically be manufactured and controlled in a laboratory setting. Parapsychologists therefore encounter coincidences after they have occurred, and must use techniques of interview and meticulous description to try to reconstruct a picture of events involved in the coincidence, much as a detective has to piece together evidence suggesting the events in a crime. Because one can never be 100% certain that all possible causal links between concurrent events have been fully investigated and eliminated, the paranormality of individual coincidences will always be a matter of degree of confidence. It is only when many coincidences are collected together and analysed that some common trends or patterns may emerge to suggest possible process-related hypotheses, some of which may be quite normal and others paranormal.

In their article 'Methods for Studying Coincidences', mathematicians Persi Diaconis and Frederick Mosteller (1989) identify 4 factors that, they feel, can account for the vast majority of coincidences. These are: hidden cause, multiple endpoints, the law of truly large numbers, and human psychology. Returning to our dictionary definition, the first factor, obviously, suggests causal connections behind coincidences; the others are related to how we find some concurrences of events more remarkable than others. The term 'human psychology' is extremely broad, however, and overlaps somewhat with the first 3 factors; after all, humans experience coincidences, so by definition human psychology is likely to play a part in all coincidences. One might refine the broad 'psychology' topic into 2 categories which are by no means mutually exclusive in the real world but which represent different schools of psychological research. The first refers to characteristics of our intuitive judgements about probability or likelihood; the second refers to ways in which our perceptions, judgements and recollections are modified so as to confirm our beliefs and expectations.

An earlier version of this paper was presented at an SPR Weekend Course on Psi and Synchronicity, November 1990; some of the other speakers focused on paranormal aspects of coincidences. I would like to thank Charles Honorton, Robert Morris and my referees for their helpful suggestions for improvements.

Unlike Diaconis and Mosteller, I am not confident that all coincidences may be explained away by these factors. Perhaps, though, an understanding of them may help parapsychologists to separate the coincidental wheat from the chaff. In this article I will briefly reiterate Diaconis and Mosteller's arguments on hidden causes, multiple endpoints, and truly large numbers, introducing other related research as we go along. I will then expand considerably upon their brief comments on psychological factors, dealing first with studies of the 'intuitive statistician' and then with ways in which our beliefs affect our perception, judgement and memory. While much of this material may already be familiar to parapsychologists, I hope to provide some service by drawing together many disparate strands of research on human judgement under uncertainty, as well as introducing some of the most recent criticisms of the 'heuristics and biases' literature.

1. Hidden Cause

Marks and Kammann (1980) described 'unseen cause' as 'the second root of coincidence' (their first root is simple probability). A coincidence is not surprising if we discover a simple reason for it. But other surprising coincidences can have perfectly straightforward hidden causes, which we have just not yet discovered. For instance, imagine a case where a woman wakes up from a nightmare in which President Gorbachev is attacked in a coup. She thinks nothing more of it, until she sees from the headlines in the following morning's newspaper that this actually happened. On first inspection this could be a meaningful coincidence, suggesting that in her dream she gained information through precognition or clairvoyance. However, when various members of the family are interviewed, it emerges that she went off early to bed the night before. The rest of the family watched the 10 o'clock news in an adjoining room, and although the woman was asleep, the news could be heard in her room. Even though she did not consciously hear the newsflash announcing the coup, this information may have been subconsciously registered, triggering the nightmare. Thus, further investigation of this coincidence between the contents of a dream and a recent news item revealed a possible hidden cause that made the coincidence less surprising. It's quite likely that a proportion of meaningful coincidences can be explained by a hidden cause. Describing the range of such causes is beyond the scope of the present article, but see Marks and Kammann (1980) and Morris (1986, 1989) for more comprehensive treatments of this topic.

2. Multiple Endpoints

A coincidence can be very impressive if it is very specific. Often, however, a 'close' coincidence is also regarded as impressive, although the chances of a 'close' coincidence happening are far greater than the chances of an exact or specific coincidence.

For example, someone may get the hunch that the phone is about to ring, and it will be Auntie Maude, who hasn't been in touch for years, making the call. As predicted, the phone does ring, only it's Auntie Maude's neighbour. Well, that's still quite an impressive coincidence, but you might also be impressed if it had been Maude's husband Bert on the line, or another auntie, or Maude's daughter...and so on.

The prediction was quite specific, but if the experient allows for 'close' coincidences to count, then the prediction has multiple endpoints. That is, there could be many 'close' coincidences that could also be seen as impressive, although the chances of a 'close' coincidence are so much higher than the chances of Auntie Maude alone being the caller.

What is it that makes a coincidence 'close'? Specific events are members of larger categories (for example, relatives who might telephone); elements in the same category or readily associated with each other (for example, a next-door neighbour of Maude) are seen in degrees of closeness in accordance with the size of the category that is shared (for example, next

door neighbour to Maude is 'closer' than a person who lives in the same town as Maude). Objectively, a close coincidence is more likely and less impressive than an exact coincidence. However, the experiencer of an exact coincidence may be nearly as greatly impressed by a close coincidence, and may even forget how specific the original prediction or hunch was, so that, with hindsight, the experiencer feels 'I knew it would happen' (a well-established phenomenon dubbed 'the hindsight effect' by Fischhoff, 1975).

In the multiple endpoints situation, we begin with a specific prediction and end with several possible outcomes. There exists a corollary to this; the so-called 'selection fallacy' (Falk, 1981-82). Scientists carry out experiments upon a random sample with pre-specified methods and analyses, and if they find the probability of their results to be very low they reject the null hypothesis and see their results as supporting the existence of a process other than chance. When an extraordinary coincidence happens, argues Falk, people often commit the logical error of singling out that one unplanned event and labelling it as significant: 'this is like the archer who first shoots an arrow and then draws the target circle around it' (Falk, 1981-82, p.25). Thus, they start with one unplanned event and proceed, *post hoc*, to regard it as significant. The event has been singled out from a range of possible other events (like categories that are shared or readily associated, as illustrated with the multiple endpoints situation) and the non-occurrence of the other possibile events is overlooked.

3. The Law of Truly Large Numbers

Events that are rare per person occur in quantity when there are large numbers of people. So, although these events are amazing coincidences to the individuals involved, they are utterly predictable if you look at the population as a whole. Of course, the precise array of events surrounding any coincidence is quite unique and can never be predicted; but it is quite predictable that something staggering is happening to someone somewhere at this precise moment. In other words 'with a large enough sample, any outrageous thing is likely to happen' (Diaconis & Mosteller, 1989, p.859).

For example, assume that daily, an incredible coincidence occurs to only one person in a million. This appears quite rare. But the population of Britain is 55 million, so each day there are likely to be 55 amazing coincidences; that makes 20,075 incredible coincidences per year. In a country such as the United States, with a population of 250 million, such incredible coincidences begin to be almost commonplace. Thus, with a large enough number of people you are bound to get amazing coincidences. It's when that statistically predictable coincidence happens to you or to someone you know that it feels spooky and you may attribute meaningfulness to it.

People are not ignorant of the fact that amazing coincidences can occur purely by chance, as only one of many possible events that could have happened. When asked to rate the surprisingness of coincidences that have happened to others, individuals are not very surprised by the accounts. When, however, they compare coincidences that have happened to themselves with those that others have experienced, the self-coincidences are consistently described as more surprising, even though others do not find these coincidences particularly surprising. Further, the more meaningful a self-coincidence is to the person involved, the more surprising it is rated as being (a 'trivial' self-coincidence might involve random numbers that are assigned to us, while a 'meaningful' self-coincidence might involve our personal names or birthdays, Falk, 1989).

This 'egocentric bias' suggests that personal involvement in a coincidence makes it seem subjectively less likely. Although we can appreciate that coincidences happening to others represent only one of a large range of possible events, when coincidences occur to us personally we do not see ourselves as 'part of the statistics'. This is a powerful effect: Falk describes how, when telling academic colleagues of the

increased surprise for self-coincidences, she was often interrupted with 'but you should hear what happened to me....' (Falk, 1989, p.488). Thus, personal involvement is one important consideration in explaining why some concurrences of events are seen as remarkable while others are not.

A similar egocentric bias may explain why, although they may be perfectly aware of the statistics for risk of death in car accidents or for risk of smoking-related disease, individuals consistently underestimate the likelihood that they personally will become victims (Slovic, Fischoff, & Lichtenstein, 1982). Experience perpetuates this myth; the newspapers only report accidents that happen to other people. It is only when someone close to us is involved in an accident or falls ill that we are suddenly reminded that we are not immune to disaster and we are not immortal!

4. Psychology

There are several aspects of human psychology that affect how we judge the likelihood and frequency of coincidences, and that affect our perception and recall of coincidences. Occasionally these psychological factors may contribute to us mistakenly judging a coincidence to be significant or meaningful.

First of all we will consider people as intuitive statisticians. I will describe the findings of research into how we make judgements under uncertainty, including estimations of likelihood or probability, and frequency or base rate information. Secondly I'll describe psychological research into how our perception, judgement and recall can be biased by our beliefs and expectations. Not all of this research has been conducted with coincidences explicitly in mind, but because the experience of coincidences is one form of judgement under uncertainty, readers may see how general psychological research may be relevant to this question.

The Intuitive Statistician

One popular illustration of how we underestimate the likelihood of a concurrence of events is the birthday problem: how many people would you need to gather together before there was a 95% chance that 2 of them would share the same day and month of birthday? The answer is surprisingly (if you are not familiar with this problem) few people; only 48 in fact. For only a 50% chance of 2 individuals' birthdays coinciding, only 23 people need be gathered together. That so few people are needed is usually quite surprising because we typically underestimate the number of different combinations of pairs of birthdays that can occur with a small number of people. We expect that with 365 possible birthdays you'd need a fairly large number of people before there was a coincidence of birthdays.

Diaconis and Mosteller (1989) have developed a simple formula that enables the calculation of the number of people needed to get a coincidence of birthdays or of any other categories: how many people (N) do you need for there to be a 50%/95% likelihood that at least 2 of them will fall in the same category from among a number of categories (c) such as 365 possible birthdates?

Approximately,

$N = 1.2\sqrt{c}$ for 50% chance

$N = 2.5\sqrt{c}$ for 95% chance

Using this formula, Table 1 shows how many people are needed for coincidences between different numbers of categories.

Table 1

Guide to solving the birthday problem, and other coincidences of categories(c)

c	=	100	200	300	*365*	400	500	600	700
N(50%)	=	12	17	21	*23*	24	27	29	31
N(95%)	=	25	35	43	*48*	50	56	61	66

PSYCHOLOGY AND COINCIDENCES

It is interesting to note how slowly N rises as c increases, so that having several hundred more partygoers does not dramatically increase the chances of a coincidence of birthdays.

Diaconis and Mosteller extend this calculation to apply to other, more complex situations, for instance, where there is more than one type of category that could coincide (such as birthdays and year of birth), and where 'close coincidences' are accepted (the multiple endpoints situation described earlier). The formula to estimate the number of people needed for a coincidence within k days in the latter, 'almost birthdays' situation (with a 50-50 chance of a coincidence) is:

$$N = 1.2\sqrt{\frac{c}{(2k+1)}}$$

With c(categories) = 365 and k = 1 day, only around 13 people are needed for a match.

These formulae may be helpful in estimating the likelihood of coincidences where the number of possible categories is known or can be discovered after some research. There remains, however, a large number of events whose frequency is difficult to measure objectively or even to estimate, and which therefore cannot be examined using such formulae. For these, as well as for coincidences that are quantifiable, people may fall back on rough 'rules of thumb'; the so-called cognitive heuristics.

Over the last 20 years cognitive psychologists, led by Amos Tversky and Daniel Kahneman, have developed the idea that people use a number of rules of thumb or cognitive shortcuts in their everyday processing of information. Usually these strategies, called cognitive heuristics, are perfectly adequate to get us through daily life efficiently. When it comes to assessing the statistical likelihood of events such as coincidences, however, it has been argued that the use of these heuristics can introduce a source of bias into our estimations. Hence, this area of research has come to be known as the 'heuristics and biases' school (e.g., Kahneman, Slovic, & Tversky, 1982; Nisbett & Ross, 1980).

As I said in the introduction, there has recently been a backlash against the heuristics and biases movement. Before I describe the reasons for this in more detail, however, I will briefly introduce 2 major heuristics (judgement by representativeness and judgement by availability) whose use may introduce some bias into people's base rate and probability estimates.

Judgement by Representativeness has been proposed to explain an apparent lack of understanding of the 'law of large numbers' (the larger the random sample, the greater its accuracy in estimating the characteristics of the parent population from which it is drawn). It is argued (e.g., Tversky & Kahneman, 1974) that people judge the likelihood of an event according to the sample's similarity to, or representativeness of, the parent population on certain essential features such as means and proportions. Sample size, which should give some indication of the degree to which one could confidently predict characteristics of the parent population, was frequently neglected by subjects in early studies by Kahneman and Tversky.

For instance, subjects were posed this question:

> 'A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50% of all babies are boys. The exact percentage of baby boys, however, varies from day to day. Sometimes it may be higher than 50%, sometimes lower. For a period of 1 year, each hospital recorded the days on which more than 60% of the babies born were boys. Which hospital do you think recorded more such days?' (Kahneman & Tversky, 1972, p.443).

Subjects' opinions were equally divided between the two hospitals, despite the fact that by the law of large numbers the smaller hospital would be expected to show more deviations from the average

50% figure. Later, however, it was demonstrated that subjects could take account of sample size if the wording of questions was simplified (e.g., Bar-Hillel, 1979); indeed, if sample size was the only information provided, then correct responding could approach 100% (Evans, 1989).

Nevertheless, in the real world, people are faced with lots of possibly irrelevant information, which may distract attention from features such as sample size that should be taken into consideration when making judgements under uncertainty. So in the case of coincidences, if people tend not to take sample size sufficiently into account when judging likelihood, they may not appreciate that an extreme outcome is more likely to occur in a small sample, and may therefore mistakenly attribute significant rarity to a coincidence occurring under these conditions.

The representativeness heuristic has also been proposed to explain the so-called 'conjunction fallacy' (Tversky & Kahneman, 1983). Here, subjects judge the conjunction of 2 events as more probable than one of its components because, it is argued, they judge according to the similarity between the paired events and an original descriptive statement; this is despite the basic tenet of probability theory that a conjunction cannot be more probable than one of its consituents. For example, subjects were given the following description (Tversky & Kahneman, 1983, p.297):

> 'Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.'

Subjects were asked to indicate which of 2 alternatives was more probable: 'Linda is a bank teller'; or, 'Linda is a bank teller and is active in the feminist movement'. 85% of the respondents indicated that the latter statement was more probably correct, a finding which Kahneman and Tversky interpret as a blatant violation of the conjunction rule.

Judgement by Availability is the second cognitive heuristic that may influence our judgements about coincidences. When we use availability we estimate frequency in terms of how easy it is to think of examples of something (Tversky & Kahneman, 1974). Like representativeness, judgement by availability is usually a good rule of thumb, but it can lead to biased decisions because availability is influenced not only by objective frequency but also by recency, familiarity and vividness. For example, when we estimate how often earthquakes occur in a 10 year period we are too heavily influenced by whether an earthquake has occurred recently.

The apparent neglect of base rate or frequency information in making probability judgements (the 'base rate fallacy') has been widely attributed to the operation of the availability heuristic (e.g., Borgida & Brekke, 1981). Here, it is argued, base rate information is often less vivid, more abstract, less noticeable than other kinds of information and so it tends to get overlooked. In the earthquake example, the base rate or frequency information refers to data about how many earthquakes have occurred in the last 10 years. Typically, this statistical information is overlooked in favour of the vivid memory of a recent earthquake, leading to an exaggerated estimation of the frequency of earthquakes. Studies that have increased the availability of base rate information (for instance by conveying it graphically rather than in tabular form) have shown that it can be taken into account by subjects.

Another consequence of the availability heuristic is that we pay less attention than we should to negative information - to non-occurrences or non-coincidences - because they are less noticeable. Logically, the failure of something to happen can be just as informative for our decision-making as a positive occurrence. Yet, because non-events are less salient or less memorable, their usefulness for judging the frequency of, say, coincidences, is neglected. Take, for example, a person who believes that she can make people telephone her simply by wishing for it to happen. When she

succeeds, and there is actually a close coincidence between the time of her act of willing and the phone ringing, this is especially memorable. Failures of the phone to ring will tend to be less noticeable. Looking back on her efforts, the lady will tend to have a higher opinion of her skills than objectively she should. This is because the successful coincidences of willing and the phone ringing will be more available than the non-coincidences, and so she is likely to overestimate the frequency of her successes. Confidence in her psychic ability may be further enhanced if she accepts close coincidences - such as phone calls coming up to 30 minutes after she willed them - as satisfactory evidence of her abilities. This is another example of multiple endpoints.

In Edinburgh we often get people calling in who think they are having psychic experiences, such as precognitive dreams. People may be wanting us to confirm their abilities, or to get rid of them. As a first step we need to get a good description of what is going on, so we may ask people to keep a diary, noting every possibly precognitive dream and whether or not it 'came true'. Doing this, people often find that they have been overestimating the frequency of their success rate, presumably because successes are so memorable. Recording actual performance can circumvent the availability heuristic to some extent, and make the non-coincidences less easy to ignore.

The heuristics and biases studies have focused on individual cognitions in relatively simple and sterile situations; subjects are often posed problems in paper and pencil form, for example. Experiments in social psychology, however, which have used more realistic or ecologically valid methodologies, have also demonstrated that availability plays an important role in our perceptions of causality. When something is available, it is more vivid, salient, noticeable, or memorable. Simple experimental manipulations of our focus of attention can dramatically influence our perceptions and causal attributions (for a review, see Taylor & Fiske, 1978), such that causality is attributed to salient stimuli.

In summary, judgement by availability may lead us to overestimate the frequency of coincidences that we expect to occur (such as predicting phone calls), and to neglect actual base rate information that conflicts with our expectations or that has low salience (for example, overlooking failed predictions).

Heuristics and Biases Re-Evaluated

Since the original influential experiments by Kahneman and Tversky which provoked a veritable flood of research, some psychologists and statisticians have begun to question the assumptions behind these studies. Criticisms have centred on: the language used to describe the effects of heuristic use; the statistical models underlying many of the studies asking subjects to make probability judgements; the methodology used to demonstrate heuristic use; and the usefulness of the cognitive heuristics in understanding judgement under uncertainty. I will cover each of these in turn, before summarising their impact on the question of coincidences.

1. Value-Laden Language. In her article 'The Rhetoric of Irrationality', Lola Lopes (1991) points out that the original heuristics and biases papers by Kahneman and Tversky focused on the *process* of judgement under uncertainty, whereas the summary article that appeared in *Science* (Tversky & Kahneman, 1974), and which therefore reached a wide audience, shifted emphasis from heuristic processing to *biased* processing. Strong evaluative language is used in this second article (e.g., 'severe errors of judgment', p.1130) and in follow-up research by other authors (e.g., 'Probability judgements are notoriously inaccurate', Blackmore & Troscianko, 1985, p.459). This language conveys a clearly critical and negative message about subjects' cognitive abilities. This might not be controversial in itself, were it not for the fact that the original experiments by Kahneman and Tversky were logically set up to identify processes rather than to

evaluate performance. Yet it is the 'inadequate intuitive statistician' message that caught the imagination and tinged the research approaches of subsequent investigators.

Lopes argues persuasively that evaluative language does not belong in scientific articles; these should be concerned with description and interpretation rather than value judgements. The 'rhetoric of irrationality' may serve to titillate authors and readers, who can feel themselves superior because (with hindsight) they can solve the probability problems; the strong language also gives the impression (misleading, as we shall see) that there is an obvious correct answer to such problems.

2. *Statistical Models.* Often the authors of papers on heuristics and biases use phrases such as 'subjects' inability to appreciate the laws of probability' or their 'lack of intuitive understanding of the normative theory of prediction'. Whereas anyone reading a standard textbook on statistics could be forgiven for concluding that there is some sort of 'normative probability theory' that provides correct answers to problems posed in some heuristics and biases experiments, those in the know - that is, statisticians - have pointed out that there *is* no normative probability theory; and, worse still, that the statistical assumptions behind the probability problems come from a school of reasoning that is held by only a minority of statisticians.

The most authoritative critic of the model of probability used in most heuristics and biases literature is Gerd Gigerenzer (e.g., 1991a, 1991b; see Gigerenzer et al., 1989, for a description of the historical development of the different statistical schools of thought; and see Gigerenzer & Murray, 1987, for a detailed consideration of these as they have been applied to the study of judgement under uncertainty). In a paper entitled 'How to make cognitive illusions disappear: Beyond heuristics and biases', Gigerenzer (1991a) makes a strong critique of the heuristics and biases school:

> What is called in the heuristics and biases literature the 'normative theory of probability' or the like is in fact a very narrow kind of neo-Bayesian view that is shared by some theoretical economists and cognitive psychologists, and to a lesser degree by practitioners in business, law, and artificial intelligence. It is *not* shared by proponents of the frequentist view of probability that dominates today's statistics departments, nor by proponents of many other views; it is not even shared by all Bayesians....By this narrow standard of 'correct' probabilistic reasoning, the most distinguished probabilists and statisticians of our century....would be guilty of 'biases' in probabilistic reasoning. (pp.86-87)

Gigerenzer proceeds to demonstrate how 'overconfidence bias' (where subjects answering a series of questions show a discrepancy between their perceived success and their actual performance of a task; overview by Lichtenstein, Fischhoff, & Phillips, 1982), the 'conjunction fallacy' and the 'base rate fallacy' can be made to 'disappear' if questions are re-phrased to take account of alternative statistical models and meanings of probability.

Let us return to the 'Linda is a bank teller and is active in the feminist movement' example used to illustrate the conjunction fallacy. Gigerenzer points out that to choose this description of Linda as more likely is a violation of *some* subjective theories of probability, including Bayesian theory, but it is not contrary to the dominant frequentist school of probability, because in this latter model, single specific events cannot be considered in terms of probability; probability theory is about frequencies, not single events. If the Linda problem is rephrased in frequentist terms 'There are 100 persons who fit the description above (i.e., Linda's). How many of them are: (a) bank tellers (b) bank tellers and active in the feminist movement' then the 'conjunction fallacy' largely disappears, with only 22% of subjects choosing option (b) as most likely (Fiedler, 1988).

Some of the 'errors' identified by Kahneman and Tversky and their followers may therefore be due to the researchers' adoption of an inappropriate statistical model rather than to weaknesses in their subjects' reasoning abilities. Further, away from the relatively controlled and clean world of the laboratory, the confusions and complexities of the real world may make the application of any statistical models controversial and rather difficult.

3. *Experimental methodology.* Earlier, when discussing the 'law of large numbers', I cited a study that demonstrated that people are more able to take account of this law if the question is phrased more simply, and if other distracting information is removed. In a similar vein, many of Kahneman and Tversky's original positions have been refined, following demonstrations that variations in experimental methodology cause variations in the appparent influence of cognitive heuristics upon problem solving and judgement under uncertainty. We have already seen how the 'conjunction fallacy' can be made to disappear by rephrasing the question.

Steven Sherman and Eric Corty (1984), for instance, review a number of studies that suggest that the extent to which heuristics are used to solve a problem may depend on the way in which the problem is presented or structured. If there is plenty of time, if the task is not too complex and is clearly presented, if base rate information is made concrete, salient and specific to an individual case, then individuals may reach the normatively correct solution (where there is one). For example, typical biases in judging random sequences can be eliminated simply by instructing subjects that random events may be present or by providing them with a comparison level of nonrandomness (Peterson, 1977).

Related to the question of experimental methodology is another telling criticism of the heuristics and biases paradigm: its lack of ecological validity. There is a considerable gulf between the sorts of paper and pencil probability problems posed to unsuspecting subjects in typical heuristics and biases experiments, and the everyday situations where judgements about probability are made (e.g., when placing a bet; when judging what caused a picture to fall off a wall; when reading about or experiencing coincidences). When such artificial situations are used in conjunction with possibly inappropriate models of probability, any conclusions that may be drawn about the use of cognitive heuristics in more complex situations become severely limited. There is a need for the heuristics and biases researchers to adopt more realistic methodologies; for instance, role-playing, simulations of complex situations, and observational studies of individuals' statistical judgements in their natural environment. As we shall see in the next section, studies of the biasing effects of beliefs and expectations on perception, judgement and memory have successfully used more realistic settings, and have produced findings that have practical applications.

4. *Theoretical usefulness of heuristics.* Sherman and Corty (1984) also note that Kahneman and Tversky's heuristics are rather vague and are often identified *post hoc*. They are insufficiently precisely defined to enable prediction of which particular heuristic will be applied in which specific situation. Gigerenzer (1991a) echoes these criticisms thus: 'All three heuristics...are largely undefined concepts and can *post hoc* be used to explain almost everything. After all, what is similar to what (representativeness), what comes into your mind (availability), and what comes first (anchoring) have long been known to be important principles of the mind' (p.102). Heuristics, he argues, are hardly more than re-descriptions of the phenomena seen in judgement under uncertainty.

Conclusions

Do these criticisms of the heuristics and biases literature negate its applicability to the question of what makes coincidences seem remarkable? Certainly, they seriously weaken those aspects of the literature that deal with probability judgements and pre-

diction where some sort of normative theory of probability has been (questionably) assumed. Further, it is difficult to generalise from the typically artificial methods used, to more complex settings. But although 'overconfidence bias' as typified in the heuristics and biases literature may 'disappear' if an alternative statistical model is adopted, in more realistic situations such as in studies of eyewitness testimony, overconfidence nevertheless remains a problem. Wells and Murray (1984), for instance, reviewed studies of eyewitnesses' confidence in their memory reports and concluded that 'the eyewitness accuracy-confidence relationship is weak under good laboratory conditions and functionally useless in forensically representative settings' (p.165).

Gigerenzer's criticisms have, however, been constructive: he suggests that the study of judgement under uncertainty may explicitly utilise various statistical models to get a clearer idea of which model most closely approximates subjects' intuitive reasoning (one might also have to consider the possibility of individual differences in model selection). Also, many statistical principles, such as the law of large numbers, are uncontroversial, and in this section I have tried to focus on aspects of judgement under uncertainty that are not so vulnerable to criticism of underlying statistical assumptions. The research on the effects of salience or availability on focus of attention and causal attributions, for example, reinforces the apparent importance of availability for judgements under uncertainty (e.g., Taylor & Fiske, 1978, Dow (Watt), 1988). Lopes' comments on evaluative language are well-taken, and are a useful reminder to all concerned with heuristics and biases that they should look out for 'creeping value judgements' in their writings.

We have seen that the degree to which heuristics are used depends greatly on the presentation of problems in the experimental situation, and that careful simplification and manipulation of information can modify or overcome heuristic use. There is no doubt, however, that in the real world, judgements have to be made under much greater uncertainty, with a profusion of distracting information and incomplete data. I believe that it is in these conditions that we are most likely to simplify by resorting to rules of thumb. If relevant information, such as base rates, is readily available and noticeable, then we have seen that it can be applied quite appropriately by individuals. On occasions when all relevant information is not at hand, heuristics may be used. Evans (1989) makes the useful distinction between competence and performance in statistical reasoning. People can be seen in some circumstances competently to apply statistical principles in judgement under uncertainty. What we need to understand is why this competence is not applied under a different set of circumstances.

The final criticism of the cognitive heuristics, that they are vague and *post hoc*, is, to me, the most telling. At the moment cognitive heuristics are largely descriptive (or heuristic!) devices to help psychologists organise their thoughts about other people's thought processes. Description is a necessary stage in the development of theoretical ideas, but the heuristics literature has yet to progress beyond this descriptive phase. We need a theory or theories of judgements under uncertainty to be developed to a stage where they offer 3 things: falsifiable predictions; an explanation of why humans judge the way they do; and predictions of the circumstances under which the various judgemental biases might be expected to operate. Describing theories of human reasoning as 'fragmented', Evans (1991) states, 'while theorists interested in bias emphasize...the role of non-reasoning processes, those interested in competence emphasize ...reasoning processes" (p.97). There is a lack of integration between the various approaches to the study of human reasoning, and Evans makes some constructive recommendations for overcoming this problem. I would agree with Sherman and Corty (1984), however, that cognitive heuristics can potentially identify the processes underlying decision-making, and can

potentially suggest how to solve decision-making problems and improve judgement. For these reasons, they may be useful in evaluating coincidences.

4.2 The Influence of Beliefs and Expectations on Perception, Judgement and Recall

Apart from characteristics of our statistical intuitions that may cause some coincidences to seem remarkable, the sense of meaningfulness of coincidences may be enhanced by other aspects of our information processing. In short, how we perceive, interpret and remember events is, to a large extent, determined by our *a priori* beliefs, expectations and theories (or schemata) about how the world works. Information that is consistent with our expectations is readily assimilated to strengthen our beliefs; on the other hand, information that does not fit with our expectations may be distorted to make it fit, selectively ignored, or forgotten, so that our prior expectations or interpretations of an event or a coincidence are not challenged.

How Beliefs Can Influence Perception and Judgement

Not only do people tend to overlook non-occurrences or their failures to get the coincidences they predicted; they also tend to see relationships where there are none. This is called the "illusory correlation" effect, and usually refers to cases where people associate 2 factors, though statistically no relationship exists. Our theories and stereotypes often lead to our perceiving illusory correlations.

The classic studies showing illusory correlation (Chapman & Chapman, 1967, 1969) were concerned with the question of why clinical psychologists persisted in reporting correlations between patients' responses on a projective psychological test, and aspects of the patients' motivations and emotions. Detailed studies of this Draw-A-Person (DAP) test suggested that responses on the test were totally unrelated to clinical symptoms. Yet clinical psychologists found that paranoid or suspicious patients exaggerated the eyes in their drawings, whereas dependent patients, who like to be fed and cared for, exaggerated the mouth. The Chapmans asked patients in a State hospital to take the DAP test. These drawings were then paired *completely at random* with 6 symptoms, such as suspiciousness and dependence. The Chapmans asked untrained college students to examine the drawings and the symptoms with which they had randomly been paired. Later, the students were asked which features of the drawings had most often been paired with each symptom. The students reported the same kinds of association between symptoms and drawings that the clinicians had, even though it had been arranged that there was no systematic relationship for the students (incidentally, these experiments do not suggest that the DAP test is of no clinical use; it may be helpful to clinicians when taken in the context of a wider clinical investigation).

Sometimes it is valid and efficient for our expectations to influence our interpretation of information; for instance, our knowledge of language may enable us to understand what is being said over a noisy telephone line. At other times, our preconceptions can be misleading; for example, where wishful thinking or preoccupation with a particular idea may lead to a misinterpretation of the caller's words. With regard to the study of coincidences, the challenge is to identify when information may have been distorted or misinterpreted. Though there is no easy answer to this problem, some pointers are given by psychological research.

Nisbett and Ross (1980) identified some factors that increase the likelihood of erroneous bias based on *a priori* beliefs or theories:

1. Confidence in the theory. If this confidence is based on emotional commitment to the theory rather than on a solid factual foundation then it is more likely that we will selectively process information so as to strengthen our beliefs.

2. *Availability of the theory.* The likelihood that a theory will influence how we interpret information depends on its availability; its likelihood of being triggered by the data at hand. If you have recently attended a course in Freudian psychoanalysis, this theory might be very available for you and be readily used to interpret the actions and dreams of people around you. A common example of the possible operation of availability in coincidences is where you learn a new word, then suddenly notice it repeatedly cropping up. It is unlikely that you have never before encountered the word; rather, your attention has been drawn to it, and it has become salient or available for you to notice when it occurs again.

3. *Ambiguity of the information.* Evidently, if information is clear and unambiguous then it may be more difficult (though not impossible) for us to put our own interpretation on that information based on our preconceptions. If, on the other hand, the information is experienced in an ambiguous way - say, in poor light, in confusing circumstances - then it is much easier for us to interpret it so as to fit our expectations. Fading of memory and the operation of our cognitive heuristics can render initially clear information ambiguous. This is why it is so important to take note of, for instance, each prediction that we make, plus whether or not it is fulfilled; and to write down details of a coincidence as soon as possible. The note-taking makes the information less ambiguous than our unassisted memory would.

How Information Often Doesn't Influence Our Beliefs

Psychological research suggests that once we have made up our minds about something we are very resistant to revising our theories. Here, I'll give examples of 3 areas of research into the effects of information on beliefs: firstly, what happens when established beliefs are faced with new information; secondly, the effect of new information on new beliefs; and thirdly, the effect of false information on beliefs.

1. *New Information and Established Beliefs.* Lord, Ross, and Lepper (1979) took 2 groups of university students: one group strongly believed that capital punishment was a deterrent to potential murderers; the other strongly believed it was worthless as a deterrent. Each subject read about the results of 2 supposedly authentic studies on the deterrent effects of capital punishment. One of the studies concluded that capital punishment was an effective deterrent. The other concluded the opposite. Subjects were asked a number of questions after they had read both studies.

There were 3 main findings from this experiment: 1. Whichever study supported a subject's own initial position was found to be significantly 'more convincing' and 'better conducted' than the study opposing their position; 2. When subjects were asked about their beliefs after reading about only one study, which could be in agreement with or in contradiction to their own views, belief in the subject's original position was strengthened if they had just read a supportive study, but belief in the original position was hardly affected at all by reading an opposing study; and, 3. After reading about *both* studies, the subjects were more convinced about the correctness of their initial position than they were before reading about *any* evidence.

In summary, different standards are used for criticizing opposing evidence to those used for criticizing supportive evidence. Mixed evidence, giving equal support to 2 opposing views, does not reduce confidence for holders of either view but instead reinforces confidence for holders of each view.

Perhaps these results were obtained because the subjects were impressionable young students. But even in the supposedly rigorous and objective world of reviewing articles for scientific journals, prior beliefs have a strong influence on evaluations. In a controversial 'real-life' experiment, Douglas Peters and Stephen Ceci (1982) resubmitted 12 already-published research

articles by authors from prestigious American psychology departments to the 12 widely-read and respected American psychology journals in which they had been originally published only 18 to 32 months earlier. The re-submitted articles were virtually identical to the originals, except that the author's name and institution were changed to fictitious ones (e.g., Dr. Wade M. Johnston at the Tri-Valley Center for Human Potential). Only 3 of the resubmissions were recognised as such. Of the remaining 9 articles, 8 were rejected on grounds such as 'serious methodological flaws'. Peters and Ceci suggest that the journal editors and reviewers may have been biased by the original authors' status and institutional affiliation. Findings such as these argue in favour of blind refereeing of academic articles.

Research along these lines provides a useful reminder that our experiences and interpretations of coincidences can be dramatically affected by our prior expectations, and that adopting an 'impartial' scientific cloak may be ineffective.

2. *New Information and New Beliefs.* Moving on now, we consider the effects of new information on new theories; specifically, the effects of the sequential processing of evidence. Usually we do not get all the information for or against a theory, or, in the context of this paper, an explanation of a coincidence, at once; some comes earlier and some comes later. People tend to base their explanations on the earliest evidence; on their first impressions. This is known as the primacy effect. Logically, when evaluating the evidence for or against a theory, all evidence is important, not just the first evidence that is encountered.

For example, in Jones et al. (1968), subjects were asked to observe another person trying to solve 30 multiple choice problems. The problems were described as being of equal difficulty. The person doing the problems always solved the same number; 15. In one condition - the descending condition - he solved a greater proportion of problems early in the series, and fewer later in the series. In the other condition - the ascending condition - he did the reverse, solving few problems to begin with, but solving more later on in the series. Subjects were asked to predict how the target person would perform with a second set of problems, to rate his intelligence, and to try to recall how many problems he had solved in the first set of problems. It was found that early performance received undue weight. The target person who solved relatively more problems early on was seen as being more likely to perform well on a second set of problems; judged as being more intelligent; and remembered as having solved more problems than the other target person.

Studies like this highlight the importance of attempting to gather all relevant information when evaluating coincidences, before forming interpretations. Of course this is difficult, especially in complex situations, but simple procedures such as using the 'birthday problem' formula to estimate the likelihood of a coincidence, may assist in a balanced evaluation of coincidences.

3. *False Information and Beliefs.* The third way in which our initial theories or beliefs fail to be influenced by objectively relevant information is when our theories persevere not in the face of new information, but rather in spite of the discrediting of the information that originally led to the formation of our beliefs or theories.

For example, Anderson (1983) presented subjects with either 2 case histories (vivid, concrete data) or a statistical summary (abstract data). The data suggested either that firefighter trainees who enjoyed risk performed well in their later careers, or that they did not perform well as firefighters. Subjects were then told that the data were fictitious. Later, it was found that they still clung to whatever initial theory they had been led to hold, and the strength of the perseverance of belief was greatest for those subjects who had seen the concrete data (this suggests the operation of salience or availability biases). This is paradoxical, because small numbers of case studies are likely to be less accurate indicators of general population characteristics than sta-

tistical summaries of a wider survey. Anderson concluded that this effect was not due to memory but to the spontaneous generation of causal explanations that seemed to be facilitated by the case histories. In the case of coincidences, of course, the data are also usually concrete; in the form of personal experiences or anecdotes that are told by others.

To sum up this section: we have seen the interplay of human psychology, beliefs, and data. We tend to cling unduly to our own beliefs or theories, even in the face of contradictory evidence, and we apply a double standard to evidence relevant to our beliefs. We have probably all seen this happening in our everyday life; but we may neglect to consider these facts when we ourselves are involved. We can easily see the weak points in other people's beliefs, while being absolutely certain of the truth of our own. This may be one reason why we are less impressed when coincidences happen to other people than when we are closely involved in them ourselves (Falk, 1989).

How Recall Can Change Due to Beliefs and Expectations

Memory is a construction, based partly on our perceptions and partly on our interpretations, and memories tend to fade and alter over time. It appears that when we recollect something we actively reconstruct our memories so as to fit with our theories and expectations. When we recall coincidences that we have heard of or have been involved in in the past, our memory may blur some details and strengthen others so as to make the coincidence seem more impressive than it was to begin with, a process which may be quite unconscious.

In 1971, Bransford and Franks developed their Constructive Model of Memory. Subjects were presented with sets of simple sentences, some of which they had seen a few minutes before and others which were new sentences, including combinations of the earlier sentences. When they were asked to identify those sentences they had seen before, many subjects were convinced they had seen the new combination sentences before. Bransford and Franks proposed that individuals integrate information from individual sentences so as to construct larger ideas; they think they have already seen these complex sentences because they have been combined in memory and, once combined, they cannot break them down into their original components.

This constructive model of memory is not necessarily limited to recall for sets of sentences. People instinctively try to make sense out of any situation - sets of noises, events happening around them, snippets of conversation - and their memories of these events may contain not only just the original events but also the interpretation put on them by the individual.

One example of the study of recollection change in more realistic situations, which are perhaps more relevant to the evaluation of coincidences, is work in the area of eyewitness testimony (e.g., Wells & Loftus, 1984). In a typical experiment, Loftus and Loftus (1975) showed subjects a film of a traffic accident. Soon after that, subjects were asked questions about their memory of the accident. One of these questions, about the speed of the cars, was asked in 2 different ways. Subjects were either asked, 'How fast were the cars going when they smashed into each other?' or they were asked 'How fast were the cars going when they hit each other?' Apparently, subjects used the different inferences suggested by the words 'smashed' or 'hit' to alter their memory of the accident. 'Smashed' implies a more destructive collision than 'hit'. A week later, subjects were given a memory test, where they were asked 'Did you see any broken glass?' Although there was no broken glass in the original film, those subjects who had been asked the 'smashed' question were more likely to say mistakenly that they remembered seeing broken glass.

The sentence-recall experiment showed how information could be misremembered only a short time after its presentation. Generally, the more time that passes after the original incident, the more chance there

PSYCHOLOGY AND COINCIDENCES

is that recollections will change. You can imagine how recall might change over months or years after an original event. This suggests that sometimes a coincidence that was only moderately impressive to begin with can, over time, be recalled differently, as really very striking.

These experiments into sentence recall and eyewitness testimony demonstrated *mis*remembering. Other studies have demonstrated *selective* remembering. Hintzman, Asher, and Stern (1978) explored their hypothesis that coincidences seem to occur more often than chance because of selective remembering of meaningfully related events, by asking subjects to rate a series of concrete nouns and, at another time, a series of pictures of objects, in a task ostensibly unrelated to memory. Some of the nouns and pictures were related to each other, but the rest were unrelated (the authors do not say by what criteria the judgements of relatedness were made). Later, participants were unexpectedly asked to recall as many words from the list of nouns as possible. This was therefore an incidental learning task, and the authors regarded the related nouns and pictures as coincidences. They found that significantly more 'related' words were recalled than 'unrelated' words, suggesting that there was selective remembering of the meaningfully-related words. An experiment of similar design but using events rather than nouns (the former being components of coincidences in the real world) replicated this selective memory retrieval effect (Kallai, 1985, cited in Falk, 1989).

In a review of the literature into 'Alterations in recollections of unusual and unexpected events', Hall, McFeaters, and Loftus (1987) described how new information could be absorbed and interpreted as an original memory. A coincidence, of course, is an unusual and unexpected event. New information might be embedded in a misleading message, or in a biasing question, or in a sketch or photograph. Private remembering of the event, discussion with friends or family, or even questioning by a careless investigator can be a source of misleading opinions and information. The experiment on eyewitness testimony described above showed how careless questioning can bias recollection.

Hall, McFeaters, and Loftus identified 4 major factors (time delay, warnings, question phrasing, and attitude) which affected the change in recollections for unusual or unexpected events. These 4 factors have been fairly well demonstrated in experiments.

The first is the *time delay* between an event, a subsequent misleading message, and a final test of recollection. It seems that changes in recollection are greatest if there is a relatively long time delay before the misinformation is given; presumably so that the original memory can fade. Then, the change in recollection is greatest if people are tested about their recall of the original information while the post-event misinformation is still relatively recent.

Secondly, it has been shown that if people are *warned* just before they are to be exposed to misinformation that the message may contain misleading information, then they are less likely to be influenced to change their original recollections. This effect is quite specific, though. If the warning is not given immediately before the post-event misinformation, then it's not usually effective.

Thirdly, it seems that the way in which a misleading question is *phrased* affects the likelihood of recollection change. After a surprise intruder interrupted their lecture, subjects who were asked, 'Was the moustache worn by the tall intruder light or dark brown?' were less likely to (mistakenly) recall that the intruder had a moustache than those who were asked' "Did the intruder who was tall and had a moustache say anything to the professor?' (Loftus, 1981). The latter question included the misinformation in an auxiliary clause, suggesting that memory is more easily altered if misinformation is casually or unintentionally absorbed, rather than being given direct and critical attention. Also, misinformation that is slowly scrutinized may be rejected, whereas if you give brief and minimal attention to the misinformation, it

may be added easily to the original recollections.

I described earlier how *attitude* can affect how we perceive or remember information. This has also been demonstrated in the experiments into eyewitness testimony. Information that is consistent with attitudes is strengthened in the process of recollection, whereas information that doesn't fit fades, or is replaced. In a classic experiment, subjects were shown a picture of 2 men in an underground train. One of the men was white, the other black. The white man held an open cut-throat razor in his hand. Subjects were asked to describe the picture to others, who in turn described it to others, and so on. It was found that, over time, the razor moved from the white man's hand to the black man's hand (Allport & Postman, 1947).

Summary and Future Directions

Some of the research described in this paper may not be new to parapsychologists, but by drawing together a variety of psychological studies relevant to the evaluation and experience of coincidences, I hope some readers may be stimulated further to consider the implications of this psychological research for the study of coincidences. I am only too aware of the limitations of this paper, which can be subjected to the same sorts of criticisms as have been levelled against the heuristics and biases approach: I have merely cobbled together a number of descriptions of relevant research findings without providing any useful explanatory framework. The various psychological factors I have described may be applied *post hoc* to account for many coincidences. What would be even more useful would be some theory or theories enabling the prediction of the circumstances under which these factors would be expected to operate. This will probably have to await further developments in mainstream psychology, though Hogarth (1981), Gigerenzer (1991a) and Evans (1991) make some constructive suggestions for how researchers could progress beyond the stage of cataloguing heuristics and biases.

A consideration of techniques for overcoming the many biases in our judgements under uncertainty would also have been helpful, but would have made the paper unacceptably long. The interested reader is referred to Kahneman and Tversky (1982), Fischhoff (1982), Nisbett et al. (1982), Evans (1989), and Lopes (1987) for further information on debiasing. Research has also been conducted into ways of improving recollection of real-world events; for example, police have an obvious interest in eyewitness recall, and Roy (1991) describes how the 'cognitive interview' has been shown to improve eyewitness recall. Four questioning strategies are used, which aim to enhance memory retrieval: the witness is encouraged to reinstate mentally the external scene and the internal thoughts that existed at the time of the crime; he or she is asked to report everything, even incomplete or apparently trivial information; events are recounted in a variety of orders; and the witness is encouraged to report events from a variety of different perspectives. The cognitive interview has been shown to facilitate retrieval of more correct information than either the standard police technique or hypnotic techniques.

In the meantime, this paper can only provide a few guidelines for coincidence research: where possible, try to get an estimation of the likelihood of a coincidence (the formulae given when discussing the birthday problem may be helpful here); search for hidden causes; guard against predictions with multiple endpoints, by documenting predictions when they are made and noting failures to confirm predictions; ask whether the interpretation of a coincidence might have been influenced by the use of representativeness and availability heuristics, especially where judgements of likelihood and causality are concerned; have several people (ideally with differing prior beliefs about coincidences) document thoroughly coincidences, to try to some extent to circumvent belief-confirming distortions in perception, judgement and memory; beware of mis-

leading a witness with carelessly phrased questions, and explore techniques (such as the 'cognitive interview') that have been successfully put to practical use by investigators in other fields.

Lest readers think that I'm being too negative about coincidences, I should stress that an understanding of the sorts of factors that can cause us to come to the wrong conclusion about coincidences, and about psi in general, works both ways. This is especially true for situations where belief plays an important role in our judgements and recall of what happened. So, just as someone who really has a strong belief that he or she is a psychic may interpret and recall events so as to back up that belief, similarly someone who's a strong goat - who strongly believes that he or she has absolutely no psychic talent whatsoever - may also interpret and recall events so as to fit with his or her belief. In that case, although something paranormal may be going on, it doesn't get acknowledged as such. Either way, an understanding of human perception, judgement and recall under uncertainty may help us, both to weed out coincidences in which no psi was involved, and to retain those which are of more direct interest to the psychical researcher.

Also, many of the points made in this paper are most relevant for the evaluation of everyday experiences of coincidences. I have not dealt with possible techniques for the deliberate quantitative assessment of coincidences, but would refer the reader to an intriguing paper by William Braud (1983) suggesting a possible methodology for such an assessment.

Looking back to my introductory paragraph, which asked what makes some concurrences of events remarkable and not others, it seems clear that beliefs, expectations and personal involvement play a large role in our amazement at coincidences. You should hear what happened to *me* the other day...

References

Allport, G.W., & Postman, L.J. (1947) *The Psychology of Rumor*. New York: Holt, Rinehart & Winston.

Anderson, C.A. (1983) Abstract and concrete data in the perseverance of social theories: When weak data lead to unshakeable beliefs. *Journal of Experimental Social Psychology*, **19**, 93-108.

Arkes, H.R., & Hammond, K.R. (Eds.) (1986) *Judgement and Decision Making: An Interdisciplinary Reader*. Cambridge: Cambridge University Press.

Bar-Hillel, M. (1979) The role of sample size in sample evaluation. *Organisational Behaviour and Human Performance*, **24**, 245-257.

Blackmore, S., & Troscianko, T. (1985) Belief in the paranormal: Probability judgements, illusory control, and the 'chance baseline shift'. *British Journal of Psychology*, **76**, 459-468.

Borgida, E., & Brekke, N. (1981) The base rate fallacy in attribution and prediction. In J.H. Harvey, W. Ickes, & R.F. Kidd (Eds.) *New Directions in Attribution Research, Vol.3*. Hillsdale, N.J.: Erlbaum.

Bransford, J.D., & Franks, J.J. (1971) Abstraction of linguistic ideas. *Cognitive Psychology*, **2**, 331-350.

Braud, W. (1983) Toward the quantitative assessment of 'meaningful coincidences'. *Parapsychology Review*, **14**, 5-10.

Chapman, L.J., & Chapman, J.P. (1967) Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, **72**, 193-204.

Chapman, L.J., & Chapman, J.P. (1969) Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, **74**, 271-280.

Diaconis, P., & Mosteller F. (1989) Methods for studying coincidences. *Journal of the American Statistical Association*, **84**, 853-861.

Dow (Watt), C.A. (1988) Factors affecting judgments about the occurrence of psi in spontaneous settings. In D.H. Weiner & R.L. Morris (Eds.) *Research in Parapsychology 1987*. Metuchen, N.J.: Scarecrow Press.

Evans, J.St.B.T. (1989) *Bias in Human Reasoning: Causes and Consequences*. Hove: Lawrence Erlbaum Associates.

Evans, J.St.B.T. (1991) Theories of human reasoning: The fragmented state of the art. *Theory & Psychology*, **1**, 83-105.

Falk, R. (1981-82) On coincidences. *The Skeptical Inquirer*, **6**, 18-31.

Falk, R. (1989) Judgment of coincidences: Mine versus yours. *American Journal of Psychology*, **102**, 477-493.

Fiedler, K. (1988) The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research*, **50**, 123-129; cited in Gigerenzer (1991a).

Fischhoff, B. (1975) Hindsight ≠ foresight: The effect of outcome knowledge on judgement under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, **1**, 288-299.

Fischhoff, B. (1982) Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.) *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Gigerenzer, G. (1991a) How to make cognitive illusions disappear: Beyond 'heuristics and biases'. In W. Stroebe & M. Hewstone (Eds.) *European Review of Social Psychology*, **2**, 83-115.

Gigerenzer, G. (1991b) From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, **98**, 254-267.

Gigerenzer, G., & Murray, D.J. (1987) *Cognition as Intuitive Statistics*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Kruger, L. (1989) *The Empire of Chance: How Probability Changed Science and Everyday Life*. Cambridge: Cambridge University Press.

Hall, D.F., McFeaters, S.J., & Loftus, E.F. (1987) Alterations in recollection of unusual and unexpected events. *Journal of Scientific Exploration*, **1**, 3-10.

Hintzman, D.L., Asher, S.J., & Stern, L.D. (1978) Incidental retrieval and memory for coincidences. In M.M. Gruneberg, P.E. Morris & R.N. Sykes (Eds.) *Practical Aspects of Memory*. London: Academic Press.

Hogarth, R.M. (1981) Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin*, **90**, 197-217.

Jones, E.E., Rock, L., Shaver, K.G., Goethals, G.R., & Ward, L.M. (1968) Pattern of performance and ability attribution: An unexpected primacy effect. *Journal of Personality and Social Psychology*, **10**, 317-340.

Kahneman, D., Slovic, P., & Tversky, A. (1982) (Eds.) *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972) Subjective probability: A judgement of representativeness. *Cognitive Psychology*, **3**, 430-454.

Kahneman, D., & Tversky, A. (1982) Intuitive prediction: Biases and corrective procedures. In D. Kahneman, P. Slovic, & A. Tversky (Eds.) *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kallai, E. (1985) *Psychological Factors that Influence the Belief in Horoscopes*. Unpublished Master's thesis, The Hebrew University, Jerusalem. Cited in Falk, R. (1989).

Lichtenstein, S., Fischhoff, B., & Phillips, L.D. (1982) Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.) *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Loftus, G.R. (1981) Mentalmorphosis: Alterations in memory produced by bonding of new information to old. In J.B. Long & A.D. Baddeley (Eds.) *Attention and Performance IX*.

Loftus, G.R., & Loftus, E.F. (1975) *Human Memory: The Processing of Information*. New York: Halsted Press.

Lopes, L. (1987) Procedural debiasing. *Acta Psychologica*, **64**, 167-185.

Lopes, L. (1991) The rhetoric of irrationality. *Theory & Psychology*, **1**, 65-82.

Lord, C., Ross, L., & Lepper, M.R. (1979) Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, **37**, 2098-2109.

Marks, D., & Kammann, R. (1980) *The Psychology of the Psychic*. Buffalo, NY: Prometheus.

Morris, R. (1986) What psi is not: The necessity for experiments. In H. Edge, R. Morris, J. Palmer, & J. Rush *Foundations of Parapsychology*. London: Routledge & Kegan Paul.

Morris, R. (1989) Parapsychology. In A. Coleman & G. Beaumont (Eds.) *Psychology Survey 7*. London: British Psychological Society and Routledge.

Nisbett, R.E., Krantz, D.H., Jepson, C., & Fong, G.T. (1982) Improving inductive inference. In D. Kahneman, P. Slovic, & A. Tversky (Eds.) *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Nisbett, R.E., & Ross, L. (1980) *Human Inference: Strategies and Shortcomings of Social Judgement*. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Peters, D.P., & Ceci, S.J. (1982) Peer-review practices of psychological journals: The fate of published articles, submitted again. *The Behavioral and Brain Sciences*, **5**, 187-255.

Peterson, C. (1977) *The Availability of the Randomness Attribution*. Unpublished manuscript, Kirkland College; cited in Sherman & Corty (1984).

Roy, D.F. (1991) Improving recall by eyewitnesses through the cognitive interview: Practical applications and implications for the Police Service. *The Psychologist*, **4**, 398-400.

Sherman, S.J., & Corty, E. (1984) Cognitive heuristics. In R.W. Wyer & T.K. Srull (Eds.) *Handbook of Social Cognition, Vol. 1*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1982) Facts versus fears: Understanding perceived risk. In D. Kahneman, P. Slovic, & A. Tversky (Eds.) *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Taylor, S.E., & Fiske, S.T. (1978) Salience, attention and attribution: Top of the head phenomena. *Advances in Experimental Social Psychology*, **11**, 249-288.

Tversky, A., & Kahneman, D. (1974) Judgement under uncertainty: heuristics and biases. *Science*, **185**, 1124-1131.

Tversky, A., & Kahneman, D. (1983) Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, **90**, 293-315.

Wells, G.L., & Loftus, E.F. (Eds.) (1984) *Eyewitness Testimony: Psychological Perspectives*. Cambridge: Cambridge University Press.

Wells, G.L., & Murray, D.M. (1984) Eyewitness confidence. In G.L. Wells & E.F. Loftus (Eds.) *Eyewitness Testimony: Psychological Perspectives*. Cambridge: Cambridge University Press.

Department of Psychology
University of Edinburgh
7 George Square
Edinburgh EH8 9JZ
SCOTLAND, UK

Psychologie et coïncidences

Résumé: Cet article présente une revue sélective de la recherche suggérant des causes normales possibles à certaines coïncidences. Après une brève discussion des causes cachées, des prédictions à issues multiples, et de simple probabilité, l'ensemble de l'article se centre sur la recherche psychologique sur le jugement et la prise de décision en situation d'incertitude. On examine les raccourcis utilisés dans le traitement d'information jugés responsables des faiblesses apparentes de nos intuitions statistiques quotidiennes, ainsi que les critiques de ce paradigme d'heuristiques et biais. On donne des exemples d'études montrant comment la perception, le jugement et le rappel peuvent être biaisés afin de confirmer nos préjugés. Certaines implications de cette recherche pour l'étude des coïncidences sont soulignées, ainsi que la recherche suggérant des mesures afin d'améliorer le jugement de façon prometteuse.

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[3]

Fantastic Memories

The Relevance of Research into Eyewitness Testimony and False Memories for Reports of Anomalous Experiences

Christopher C. French

Abstract: *Reports of anomalous experiences are to be found in all known societies, both historically and geographically. If these reports were accurate, they would constitute powerful evidence for the existence of paranormal forces. However, research into the fallibility of human memory suggests that we should be cautious in accepting such reports at face value. Experimental research has shown that eyewitness testimony is unreliable, including eyewitness testimony for anomalous events. The present paper also reviews recent research into susceptibility to false memories and considers the relevance of such work for assessing reports of anomalous events. It is noted that a number of psychological variables that have been shown to correlate with susceptibility to false memories (e.g., hypnotic susceptibility, tendency to dissociate) also correlate with the tendency to report paranormal and related anomalous experiences. Although attempts to show a direct link between tendency to report anomalous experiences and susceptibility to false memories have had only limited success to date, this may reflect the use of inappropriate measures.*

I: Introduction

In all cultures throughout the world, there have always been occasional reports of strange, even miraculous, events. Today, such events are often labelled as 'paranormal' to indicate that, if they really did occur as reported, conventional science is incapable of explaining them. Such reports have always aroused intense controversy. Believers in the paranormal see them as proof of the limitations of the scientific worldview, whereas sceptics often dismiss them as being the result of fraud, stupidity or madness. How should a fair-minded, intelligent, rational person respond to such reports?

Correspondence:
Christopher C. French, Anomalistic Psychology Research Unit, Dept of Psychology, Goldsmiths College, University of London, New Cross, London SE14 6NW, UK. *Email: psa01ccf@gold.ac.uk*

In 1748, Hume published *Of Miracles*, an essay that is particularly relevant to this question (Grey, 1994). Hume presented a strong argument that one would never be rationally justified in believing that a miracle had occurred. He defined a miracle as an event that violates a law of nature, a definition that would be taken by many as including paranormal events. It is important to realise that Hume was not claiming to have proved that miracles have never occurred, only that we would never be justified in believing that they have. He proposed the following principle:

> No testimony is sufficient to establish a miracle unless that testimony be of such a kind that its falsehood would be more miraculous than the fact which it endeavours to establish.

Although this principle allows for the possibility that the evidence in favour of a miracle might outweigh the evidence against it, in practice, Hume argued, this never happens. Several factors undermine the credibility of miraculous claims, not least of which is the problem of witness reliability. Is it more likely that the person or persons making the claim are deceivers, or else themselves deceived, or that a law of nature has been violated? Whereas the evidence supporting violations of laws of nature is sparse, possibly even non-existent, we are surrounded by evidence that people sometimes lie and sometimes make mistakes.

This article will focus on the reliability of accounts of anomalous events from individuals who are sincere in presenting those accounts. This is not to deny that deliberate hoaxes and fraud are present in the world of the paranormal, but to accept that many — probably most — reports are made in good faith. Even so, sincerity is no guarantee of accuracy. Empirical support for such scepticism comes from both classic experiments on eyewitness testimony and more recent research on the formation of false memories. The evidence will be considered under four headings. (1) Cases where a normal episode is generally agreed to have taken place, but eyewitnesses disagree over details of what happened. (2) Cases where an apparently paranormal episode is generally agreed to have taken place, but eyewitnesses disagree over details of what happened. (3) Cases where there is a doubt as to whether a sincerely remembered normal episode ever took place at all. (4) Cases where a sincerely 'remembered' episode can be shown never to have taken place, but is entirely the product of an experimental procedure of one kind or another.

In the light of this survey, and accompanying analysis of the ways in which sincere memories can be inaccurate, the role of unreliable memory as a source of genuinely held belief in paranormal events will be considered. The motivation for this investigation is that sceptics are often challenged to offer a natural explanation for some alleged paranormal event as described by an individual who claims to have witnessed the event firsthand. However, if memory research supports the idea that such an account may not be an accurate reflection of an actual past event, then in the absence of objective evidence that the event really did occur as described, this is a challenge that should not be accepted. It is possible that the 'event' is either a distorted account of an episode that did occur or even a false memory with no basis whatsoever in objective reality.

II: Evidence of Unreliable Testimony

1. Eyewitness testimony (normal events)

A great deal of research has been directed towards understanding factors affecting the reliability of eyewitness testimony for normal events, particularly in forensic contexts. A full review of this topic is beyond the scope of the current paper (for more detailed consideration, the reader is referred to, e.g., Cohen, 1989; Loftus, 1979). Studies have typically involved assessing the recall of eyewitnesses for staged events, either using live action or video presentation. When we are able to assess witness reports against some form of objective record, it becomes clear that both perception and memory are constructive processes, influenced not only by input from the senses ('bottom-up' influences) but by our own knowledge, belief and expectations about the world ('top-down' influences).

Some of the findings from this body of research are consistent with our everyday intuitions about how memory works. Our memory is less accurate for peripheral details compared to those upon which our attention is focussed. Our memories are poorer for stimuli presented briefly under imperfect viewing conditions compared to extended viewing under ideal conditions. Our memories are most accurate when we are neither under-aroused (e.g., drowsy) nor over-aroused (e.g., frightened).

However, it is worth noting that our intuitions about memory are often wrong. For example, even under perfect viewing conditions, our memories of what we saw may be highly influenced by our view of what we think we must have seen. French and Richards (1993) showed participants an ordinary clock face with Roman numerals under perfect viewing conditions for an extended period. Participants were asked to draw the clock face from memory. They tended to represent the four as 'IV' in line with their general expectations of Roman numerals. In fact, however, the four on clocks and watches is almost always represented as 'IIII'. Most people are quite surprised when this is first pointed out to them, as they reflect upon the literally thousands of occasions they must have looked at clocks and watches without noticing this oddity. Even thousands of exposures to a simple stimulus under perfect viewing conditions may not be enough to lead to accurate recall.

Schema theory provides a useful framework for considering eyewitness testimony. As Cohen (1989, p. 71) points out:

> It can account for the fact that many of our experiences are forgotten, or are reconstructed in a way that is incomplete, inaccurate, generalised, or distorted. Schema theory emphasises the role of prior knowledge and past experience, claiming that what we remember is influenced by that which we already know. According to this theory, the knowledge we have stored in memory is organised as a set of schemas, or knowledge structures, which represent the general knowledge about objects, situations, events, or actions that has been acquired from past experience.

Loftus (1979) has drawn attention to the potentially distorting effects on memory of the use of leading questions in post-event interviews. Classic examples from

her research include the fact that, following the viewing of a film of two cars colliding, witnesses give much higher estimates of speed if they are asked how fast they were travelling when they 'smashed into' each other than if the word 'contacted' is used. Furthermore, witnesses were more likely to report seeing a broken headlight (even though there wasn't one) if they were asked, 'Did you see *the* broken headlight?' as opposed to 'Did you see *a* broken headlight?' (emphasis added). There are real concerns that interviews carried out by investigators with very strong motivations to find evidence supporting their beliefs may often unintentionally lead witnesses in similar fashion.

Another potential source of socially encountered misinformation is that of fellow eyewitnesses. Understandably, investigators often have more faith in an eyewitness account if it appears to be supported by an account of the same incident from another eyewitness. However, it is very likely to be the case that witnesses will have discussed the incident amongst themselves before ever being formally interviewed by investigators. In the light of findings from research on conformity, we might expect that witnesses will influence each other's reports to a greater or lesser extent. Recent experimental work (e.g., Gabbert *et al.*, in press, submitted) has shown that this is indeed the case. In a sense, such research on misinformation effects provides a link between that dealing mainly with naturally arising memory distortions for witnessed events and that dealing primarily with false memories for events that never actually took place at all.

2. Eyewitness testimony (apparently paranormal events)

The focus of this section is on the reliability of eyewitness accounts of apparently paranormal events in circumstances where we can be fairly certain that paranormal forces were not at work. In many of the studies reviewed below, we can be absolutely certain of this, as the situations employed were entirely under the control of the investigators concerned.

As long ago as 1887, Davey had experimentally demonstrated the unreliability of eyewitness accounts of séances. Hyman (1985, p. 27) offers the following account:

> Davey had been converted to a belief in spiritualistic phenomena by the slate-writing demonstrations of the medium Henry Slade. Subsequently, Davey accidentally discovered that Slade had employed trickery to produce some of the phenomena. Davey practised until he felt he could accomplish all of Slade's feats by trickery and misdirection. He then conducted his well-rehearsed séance for several groups of sitters, including many who had witnessed and testified to the reality of spiritualistic phenomena. Immediately after each séance, Davey had the sitters write out in detail all that they could remember having happened during his séance. The findings were striking and very disturbing to believers. No one realized that Davey was employing tricks. Sitters consistently omitted crucial details, added others, changed the order of events, and otherwise supplied reports that would make it impossible for any reader to account for what was described by normal means.

Similar findings were reported by Besterman (1932) and, more recently, by Wiseman *et al.* (1995) and Greening (2002).

Such studies do not allow us to distinguish between the possibilities that distortions occurred during the actual perception of the events as opposed to subsequent recall, but the end result is the same. It is likely that both stages of information processing are affected. The actual perception of the séance is likely to be affected by the mental set that is adopted by the observer. A sceptic is likely to adopt a 'problem-solving' approach, intent on trying to figure out how any effects are being achieved. Someone who believes that the effects might be genuinely paranormal is more likely to just sit back and enjoy them, without a critical eye for crucial details.

Such an account is supported by Wiseman and Morris (1995) who compared believers and disbelievers in the paranormal in terms of their accuracy of recall for pre-recorded 'pseudo-psychic demonstrations' — in other words, conjuring tricks, such as bending a key ostensibly using psychic powers. Overall, believers tended to rate the demonstrations as being more 'paranormal' than disbelievers. They also tended to be less accurate in remembering information that was crucial to explaining how the deception had occurred (e.g., the fact that the key disappeared from view was important because a bent key was switched for the original straight key).

Jones and Russell (1980) exposed participants to either a 'successful' demonstration of ESP or a 'failed' demonstration. In the former case, the experimenters used a marked deck of cards to ensure 60 per cent accuracy, whereas performance was at the chance level of 20 per cent in the latter demonstration. Results again showed accurate recall by disbelievers regardless of whether the results supported their belief, but a strong tendency for believers to remember both demonstrations as successful.

Many of the factors associated with poor reliability of eyewitness testimony are commonly (although not universally) associated with ostensibly paranormal events (see, e.g., Loftus, 1979). These include poor viewing conditions (e.g., darkness or semi-darkness), altered states of consciousness (e.g., due to tiredness, biological trauma, engaging in particular rituals or drug abuse), emotional arousal, and either the ambiguous and unexpected nature of the event on the one hand (in spontaneous cases) or a high level of expectation and will to believe on the other (e.g., in a séance). It should not be surprising, therefore, that the often schema-driven accounts given by eyewitnesses to ostensibly paranormal events are typically distorted versions of the actual events in question. French (1992; 2001a) discusses further the role of beliefs and expectations in perception and interpretation of anomalous experiences.

3. Questionable normal memories

Loftus (1993) presents evidence showing that autobiographical memories for entire episodes can often be open to doubt, even in the absence of any deliberate attempt by others to implant such memories. For example, she describes a study by Pynoos and Nader (1989) in which the investigators had assessed children's memories for a sniper attack on an elementary school playground. Interestingly,

children who had not, in fact, been present during the attack provided apparently sincere first-hand accounts of the event, presumably based upon accounts provided by actual witnesses:

> One girl initially said that she was at the school gate nearest the sniper when the shooting began. In truth she was not only out of the line of fire, she was half a block away. A boy who had been away on vacation said that he had been on his way to the school, had seen someone lying on the ground, had heard the shots, and then turned back. In actuality, a police barricade prevented anyone from approaching the block around the school. (Pynoos & Nader, 1989, p. 238)

Another relevant example is provided by so-called 'flashbulb memories' (Brown & Kulik, 1977). It was once believed that certain highly emotional events could lead to memories that were highly vivid and accurate. Classic examples include people's highly confident reports of where they were, whom they were with and what they were doing when they learned of some dramatic news story, such the assassination of John F. Kennedy. Subsequent research in which participants were questioned soon after such dramatic events, and then again after a long delay, has shown that even flashbulb memories can often be inaccurate, no matter how confidently they are described (see, e.g., Neisser & Harsch, 1993, pp. 9–31, for a study of flashbulb memories of the Challenger disaster).

Loftus (1993) provides numerous other examples of situations where confidently held autobiographical 'memories' appear to be based upon no event that the claimant ever actually witnessed first-hand. Such examples should lead us all to be somewhat less confident concerning the accuracy of our autobiographical memories, no matter how clear and vivid they may appear to be. There are very few real-life contexts in which we are forced to question the accuracy of apparent memories, either our own or those of others. But it appears likely that many such memories, whether for natural or paranormal events, could be false memories even without any deliberate attempt by others to implant such a false memory.

4. Implanted false memories

Although psychologists have long recognised that eyewitness accounts of witnessed events were unreliable, it is only within the last decade or so that much research has been directed at the possibility that people may sometimes have rich and detailed memories for events that they have never actually witnessed at all. The main reason for this explosion of research into false memories was the sudden increase in cases of alleged recovered memories of childhood sexual abuse, especially in the USA (see, e.g., Lindsay & Read, 1995; Loftus, 1993; Loftus & Ketchum, 1994; Ofshe & Watters, 1994). Worryingly, surveys of some professionals who were using such techniques as hypnotic regression in attempts to recover memories of abuse revealed a very poor understanding of the relationship between hypnosis and memory. In Yapko's (1994, p. 163) words:

> Survey data regarding hypnosis and suggestibility indicate that while psychotherapists largely view hypnosis favourably, they often do so on the basis of misinformation. A significant number of psychotherapists erroneously believe, for example,

> that memories obtained through hypnosis are more likely to be accurate than those simply recalled, and that hypnosis can be used to recover accurate memories even from as far back as birth. Such misinformed views can lead to misapplications of hypnosis when attempting to actively recover memories of presumably repressed episodes of abuse, possibly resulting in the recovery of suggested rather than actual memories.

Experimental psychologists tended to doubt the accuracy of the memories recovered via hypnosis and related techniques (e.g., Spanos, 1996; Wagstaff, 1989, pp. 340–57). A considerable amount of experimental evidence shows that the hypnotic regression procedure is such that it provides a context in which individuals often produce an account mixing fantasy with pre-existing knowledge and expectations — and may then come to believe with total conviction that the account reflects events that really took place (McConkey *et al.*, 1998, pp. 227–59). Recent reviews by Kebbell and Wagstaff (1998) and Lynn and McConkey (1998) conclude that there is little or no evidence to support the claim that hypnosis can reliably enhance the accuracy of eyewitness memory.

Indeed, experimental psychologists have expressed doubts about the very concept of repression itself. The idea that the unconscious mind can somehow automatically take over and hide away memories for traumatic events is not supported by any convincing experimental evidence (Holmes, 1990, pp. 85–102). However, it must also be recognised that convincing experimental evidence for repression would be almost impossible to produce for ethical reasons. The severity of the traumatic intervention which clinicians suspect would be required to produce repression is far more extreme than the experimental manipulations that any ethics committee would approve.

Data are available from real-life contexts supporting the claim that some people experience traumatic events and subsequently appear to be unable to recall those events. Williams (1994), for example, showed that many women with a documented history of childhood sexual abuse did not report the abuse when interviewed twenty years later. Loftus *et al.* (1994) reported that 19 per cent of their sample of women reporting childhood sexual abuse felt that they had forgotten the abuse for periods of their life, only for the memory to return later. There are numerous difficulties in interpreting the findings from such studies as they relate to the concept of repression. Loftus *et al.* (1994) consider a number of these, including the fact that some such events would elude recall due to childhood amnesia and, in other cases, the ordinary mechanisms of forgetting. Furthermore, it is possible that some women may actually remember the abuse but choose not to reveal this to the interviewer. Femina *et al.* (1990), in a study of childhood physical abuse, found that some interviewees with documented abuse histories simply denied ever having been abused. However, when confronted with the evidence of abuse during a second follow-up interview, the interviewees admitted they could remember the abuse. Reasons for initially denying the abuse included 'embarrassment, a wish to protect parents, a sense of having deserved the abuse, a conscious wish to forget the past, and a lack of rapport with the interviewer' (p. 229).

To a large extent, whether or not repression ever occurs, in the sense of an active, unconscious, automatic and involuntary suppression of traumatic memories, is not centrally important to the issues addressed in this paper. Readers are referred to collections edited by Conway (1997), Davies and Dalgleish (2001), Lynn and McConkey (1998), Pope and Brown (1996) and Schacter (1995), for a range of views on the wider issues surrounding this debate. For our purposes, it is sufficient that the controversy led to increased research activity in the area of false memories.

In the early days of the controversy, those who believed that recovered memories were largely accurate would sometimes object that, although memory for peripheral details of a witnessed event might be distorted, there was little evidence that people were prone to false memories for episodes that had never actually occurred at all. In fact, we now know that it is alarmingly easy to implant false memories in a sizeable minority of the population using well-established experimental techniques.

There is now a considerable amount of experimental literature available regarding false memories. However, it is unclear whether different experimental approaches lead to different types of false memory and as yet no single theory can fully account for all of the available data. Intuitively, some experimental approaches appear to be of greater relevance to assessing the likely reliability of accounts of anomalous events than others. The different approaches described below vary in a number of important ways. Some approaches involve studying distortions of memory for events that were actually witnessed, whereas more recently attempts have been made to implant false memories for entire episodes that were never witnessed at all.

Some commentators would include the extensive literature on the so-called 'misinformation effect' established by Loftus and colleagues in the 1970s (e.g., Loftus *et al.*, 1978) within the false memory framework. In general, such studies have involved showing participants slides or video clips of events such as traffic accidents or criminal acts and subsequently reading text which includes misleading information about the witnessed event. Participants frequently incorporate the misinformation into their memory for the event as demonstrated in recall or recognition tasks. Studies investigating the effects of leading questions upon recall and recognition (e.g., Loftus, 1975), as described above, can also be considered within the misinformation framework insofar as they involve a deliberate attempt to distort a memory for an actual witnessed event. The actual events in question may range from the trivial (e.g., falsely recalling single words) to the mildly traumatic (e.g., getting lost as a child in a shopping mall).

Roediger and McDermott (1995) replicated an effect first demonstrated by Deese (1959) in which participants were presented with a list of words all strongly semantically related to a critical non-presented word. For example, the words *thread, pin, sewing, point,* and so on, were presented, but the word *needle* was not. Subsequently, the critical lure word (in this example, *needle*) was falsely recalled or recognised with great confidence by many participants.

Arguably of more direct relevance to reports of anomalous events are those studies that have attempted to implant false memories for entire episodes that in all probability never occurred. Loftus and Pickrell (1995), for example, found that partial or complete false memories for a plausible but false childhood event (i.e., getting lost in a shopping mall) could be implanted in around 25 per cent of their participants. This was achieved by repeatedly interviewing the participants and getting them to try to recall as much detail as possible for four childhood events, three of which had actually occurred (according to other family members) plus the false event. Similar results were reported by Hyman *et al.* (1995). Other techniques that have been successfully used to implant false memories rely upon the use of other forms of false feedback to convince participants that events that they initially cannot remember must have actually taken place. Mazzoni and Loftus (1998) found that telling participants that the contents of their dreams indicated that certain events must have taken place before the age of three led to a dramatic increase in reports of corresponding memories.

Orne (1979) is one of a large number of investigators to show that suggestions made to hypnotically susceptible individuals following a hypnotic induction procedure will often lead to those individuals reporting memories for events that never occurred (e.g., being woken up in the night by a loud noise). However, numerous studies have now demonstrated that simply imagining events that never occurred can also lead to the formation of false autobiographical memories, a phenomenon that is known as 'imagination inflation' (Loftus, 2001). For example, Garry *et al.* (1996) had participants indicate which of a number of childhood events had or had not happened to them personally. Two weeks later, participants were asked to imagine some of the events that they had indicated had not happened to them. Subsequently, their confidence that these events had actually taken place was significantly increased relative to similar events that had not been imagined.

Further research is needed on the relationship between laboratory-based measures of susceptibility to false memories and susceptibility to false memories in more ecologically valid contexts. In general, experimenters adopt a single measure of false memory formation and so it is not clear whether or not the different measures would all inter-correlate, supporting the notion of a general susceptibility to false memories. Alternatively, it may be more accurate to think in terms of a number of distinct susceptibilities, each of which are related to different underlying brain mechanisms. It is worth noting, however, that Platt *et al.* (1998) reported that susceptibility as assessed using Roediger and McDermott's (1995) word-list technique was positively correlated with susceptibility to false autobiographical memories.

A full discussion of different theoretical approaches to false memories is beyond the scope of the current article (see Brainerd *et al.*, 2000, pp. 93–127, for further details). Although some apparent false memories can be accounted for in terms of demand characteristics and participants actually reporting accurate memories for misinformation (McCloskey & Zaragoza, 1985), it is now generally accepted that false memories really can be produced using the paradigms

described above. Early 'single-trace' theories assumed that only one memory trace was laid down for each event and that this trace had to be overwritten or distorted in some way for a false memory to result. However, such theories have largely been supplanted by 'multiple-trace' theories in which more than one memory trace is associated with each event and false memories occur when there is confusion regarding which traces are accurate.

One influential theory of false memory development is that put forward by Hyman and Kleinknecht (1999, pp. 175–88). They proposed that three processes are involved in the development of false memories. First, the presented information is judged with respect to plausibility. Such judgements will be dependent upon the source of the information and the pre-existing beliefs of the individual. Second, an event memory must be constructed on the basis of schematic knowledge plus personal experiences, suggestion and current situational demands. Finally, the individual must commit a source monitoring error in which the constructed memory is accepted as reflecting the initial event rather than misinformation presented following the event.

Several other models of false memory also assume that errors of source monitoring underlie false memories. Source monitoring refers to the ability to accurately determine the original source of information (Johnson *et al.*, 1993) where the original sources could represent any number of internal or external sources. Internal sources might include imagination, dreams or hallucinations. External sources might include written text, pictures, verbal utterances (by a range of speakers) and so on. One particular aspect of source monitoring which is of potentially great relevance to the topic of this paper is that of reality monitoring, i.e., the more general ability to distinguish between memories based upon external events and those generated by internal mental processes (Johnson & Raye, 1981). Indeed, a number of experimental techniques have been developed to allow measurement of reality monitoring ability in which participants are presented with some stimuli and asked to internally generate others, for example by imagining them. The number of errors made in subsequently deciding which stimuli were presented and which were simply imagined provides an index of reality monitoring ability. Several commentators view errors in which imagined items are confused with presented items as themselves being false memories and have used standard reality monitoring tasks to assess susceptibility to false memories (e.g., Blackmore & Rose, 1997).

III: Further Examples of Probable False Memories for Anomalous Events

Section II.2 above dealt with several instances where one could be certain that the situations concerned did not involve paranormal forces because they were entirely under experimental control and objective records existed of the events involved. It is often the case, however, that the only source of evidence is the allegedly first-hand report itself. In the case of alien abduction claims and past-life regressions, a strong circumstantial case can be made that we are indeed often dealing with instances of false memory.

1. Alien abduction claims

It appears that the circumstances under which detailed reports of alien abduction are produced are exactly those that one would expect to lead to the formation of false memories. Blackmore (1994, p. 30) provides the following report of an alleged alien abduction. It is a fictional composite account based upon her investigations of numerous cases:

> I woke up in the middle of the night and everything looked odd and strangely lit. At the end of my bed was a four-feet-high grey alien. Its spindly, thin body supported a huge head with two enormous, slanted, liquid black eyes. It compelled me, telepathically, to follow and led me into a spaceship, along curved corridors to an examination room full of tables on which people lay. I was forced to lie down while they painfully examined me, extracted ova (or sperm) and implanted something in my nose. I could see jars containing half-human, half-alien foetuses and a nursery full of silent, sickly children. When I eventually found myself back in bed, several hours had gone by.

Those who believe that alien abduction accounts accurately reflect events that really occurred often argue that the aliens involved are generally capable of rendering their victims almost totally amnesic with respect to the episode. The abductee may, for example, only remember waking up in his or her bedroom and being unable to move. Alternatively, the abductee may remember nothing at all, and simply be aware of a period of 'missing time'. Such experiences are open to various more conventional explanations, but some ufologists (e.g., Hopkins *et al.*, 1992) insist that they actually indicate a high probability that the victim was abducted. If such an explanation appears to offer a plausible explanation to the person who experienced it, that person may be interviewed by a therapist specialising in alien abduction cases to see if they can recover further details. Spanos *et al.* (1994, p. 438; see also, Spanos, 1996) comment as follows:

> Frequently, the interviews include two phases. In the first phase background information is obtained and clients are asked about unusual or inexplicable experiences that have occurred during their life. These include 'missing time' experiences, unusual or bizarre dreams, and experiences that suggest hypnagogic imagery or sleep paralysis (e.g., having seen a ghost, strange lights, or a monster). Such experiences are defined as distorted memories of alien abduction that call for further probing (Jacobs & Hopkins, 1992). Moreover, making such experiences salient enhances the likelihood that some of their characteristics (e.g., paralysis, feelings of suffocation) will be incorporated into any abduction memories that are recalled in Phase 2. Phase 2 typically involves hypnotic or non-hypnotic guided imagery employed to facilitate recall. This may involve leading questions (Baker, 1992), or the subject may be pressed repeatedly for more details (Jacobs, 1992). In addition, subjects may be informed that some material is so deeply hidden that several such interviews are required. Subjects who have difficulty 'remembering' some or all of their abduction are defined as 'blocking' and are provided with strategies for facilitating recall. These include asking subjects to imagine a curtain and then to peek behind it to view their abduction, or to imagine a movie screen on which they see their abduction replayed (Jacobs & Hopkins, 1992).

The creation of false memories is clearly implicated in UFO abduction claims, but several other factors are also involved (see, e.g., Appelle *et al.*, 2000, pp. 253–82; French, 2001b, pp. 102–16; Holden & French, 2002), although discussion of such additional factors is beyond the scope of this article.

2. Hypnotic past-life regression

Some believers in reincarnation believe that it is possible to hypnotically regress individuals not only back to childhood, but back to previous incarnations. A Cardiff-based hypnotherapist, Arnall Bloxham, was the subject of a BBC documentary and subsequently featured in a book by Iverson (1977). At first sight, it appeared that Bloxham had used hypnotic regression to produce incontrovertible proof of reincarnation. One of his cases in particular, that of a Welsh housewife referred to as Jane Evans, appeared to be very impressive. She provided details of six previous incarnations, many with a wealth of historically correct background information. In one life, she was a maid in the house of a wealthy French merchant named Jacques Couer in the fifteenth century.

Although Iverson felt that the case for reincarnation was established, subsequent investigation by Harris (1986) proved him wrong. In fact, in both of these cases and others, there were significant errors in the accounts produced. For example, Jane Evans reported that Couer was single with no children. In fact, he was married with five children — something that most maids would notice. Such errors provided the clue to the source of the story. A novel by Thomas B. Costain entitled *The Moneyman* was based upon Couer's life but the author had taken the literary liberty of deliberately omitting Couer's family as they kept getting in the way of the plot development. It appears that Evans had read the book and then forgotten reading it. During the hypnotic sessions these details had re-emerged and had been taken to be real memories.

In the case of Jane Evans and many other similar claims, it is generally believed that no deliberate hoax was involved. Instead, these are seen as being cases of cryptomnesia (literally, 'hidden memories'; see Baker, 1992). It is argued that an individual can store away information from a variety of sources during his or her life, such as from novels, films, history books, or wherever, without later being aware of the source of the information. When the information is later recalled under hypnosis, perhaps elaborated upon by the individual's own fantasies, the memories can be taken to be veridical.

Spanos and colleagues (1994) summarise some of their own studies of past-life regression. It appears that a particular type of personality is very prone to producing detailed accounts of past lives under hypnosis. Such individuals score highly on measures of fantasy-proneness. They are highly imaginative individuals with a rich fantasy life and sometimes have difficulty separating fantasy from reality. They become engrossed in works of fiction to the extent that they lose themselves. Elsewhere, Spanos *et al.* (1991) have reported the results of studies in which individuals were hypnotised and regressed into past lives and then asked for details of their past life. Information that any individual living at

the time would be aware of (e.g., the country's currency, ruler, etc; is the country at war?) is usually not known by the participant. Whether or not participants subsequently accept their past-life memories as evidence of reincarnation depends upon whether they believe in the possibility of reincarnation and the expectations built up by the experimenter.

IV: Is There a Link Between Susceptibility to False Memories, Paranormal Beliefs and Tendency to Report Paranormal Experiences?

Within the last few years, attention has turned to the issue of why some individuals seem to be more prone to false memories than others. A number of psychological factors have been identified as being correlated with such susceptibility and the degree to which such factors have been found to correlate with the tendency to report anomalous experiences is of considerable interest. If common factors were found linking both susceptibility to false memories and tendency to report anomalous experiences, this would strengthen the *prima facie* case that at least some reports of anomalous experiences may be based upon false memories. Not surprisingly, the link between paranormal belief and reports of personal experience of ostensibly paranormal phenomena is already well established. Those who feel they have had personal experience of the paranormal are understandably far more likely to believe in the paranormal.

Dobson and Markham (1993) and Markham and Hynes (1993) reported that participants with vivid visual imagery were more likely to make source-monitoring errors. Hypnotic suggestibility has been found to correlate with number of false memories reported by a number of investigators (e.g., Barnier & McConkey, 1992; Laurence & Perry, 1983; Sheehan *et al.*, 1991). Heaps and Nash (1999) found that susceptibility to imagination inflation was correlated with indices of hypnotic suggestibility and dissociativity, but not with vividness of imagery or interrogative suggestibility. However, a subsequent study by Horselenberg *et al.* (2000) did find a correlation between imagination inflation and imagery ability, using a different measure of the latter. Tomes and Katz (1997) assessed habitual susceptibility to misinformation by presenting participants with three events involving misinformation. They found it to correlate with vivid visual imagery (as well as spatial dexterity and emotional empathy for others). Eisen and Carlson (1998) reported that susceptibility to misinformation was positively correlated with both absorption and dissociation. *Absorption* has been described by Tellegen and Atkinson (1974) as 'a disposition for having episodes of single "total" attention that fully engage one's representational (i.e., perceptual, enactive, imaginative and ideational) resources'. It is commonly measured using the Tellegen Absorption Scale (Tellegen & Atkinson, 1974).

Hyman and Billings (1998) attempted to implant false childhood memories in participants using a similar approach to that employed by Loftus and Pickrell (1995). Using this technique, susceptibility to false memories was found to correlate with scores on the Creative Imagination Scale (CIS; a measure of both hypnotisability and imagery) and dissociativity, but not with absorption or social

desirability. Platt *et al.* (1998) used two measures of memory distortion in their study: scores on Roediger and McDermott's (1995) word task (described above) and naturally occurring distortions of autobiographical memory. Only one significant correlation was found between either of the two memory measures and measures of absorption, dissociativity and fantasy-proneness: absorption was negatively correlated with accuracy of autobiographical memory. Although no significant correlations were found between the word task and personality measures in this study, a previous investigation by Winograd *et al.* (1998) had found significant correlations between both dissociativity and vivid imagery and susceptibility to false memories. CIS scores were not correlated with proneness to false memories in this study.

A number of psychological factors have thus been found to correlate with susceptibility to false memories, although there is considerable variation across studies. It is likely that this reflects, to some extent, the use of different measures of susceptibility, suggesting that different types of false memory may well depend upon different mechanisms. As stated, if the same psychological factors also correlate with paranormal belief and reports of anomalous experiences, it would strengthen the argument that at least some such reports may depend upon false memories.

Imagery ability has also been found to correlate with paranormal beliefs (Finch, 2002; Greening, 2002; Diamond & Taft, 1975). Furthermore, although people who report out-of-body experiences (OBEs) do not score higher than non-OBEers on standard imagery questionnaires (e.g., Blackmore, 1982; Irwin, 1981a), they do seem to be superior in terms of using spatial imagery to create novel perspectives (e.g., Blackmore, 1986, pp. 108–11; Cook & Irwin, 1983).

A number of studies have demonstrated a small but significant correlation between hypnotic susceptibility and belief in the paranormal (e.g., Diamond & Taft, 1975; Palmer & Van Der Velden, 1983; Wagner & Ratzenberg, 1987), although some studies have failed to find such a relationship (e.g., Groth-Marnat *et al.*, 1998–99; Pekala *et al.*, 1995). Atkinson's (1994) study is exceptional in finding a relatively large correlation ($r = .53$) between hypnotic susceptibility and belief in the paranormal. Other investigators have shown that groups of highly hypnotisable participants report higher levels of paranormal belief than those with less susceptibility (Nadon *et al.*, 1987; Pekala *et al.*, 1992; Pekala *et al.*, 1995; see Kumar & Pekala, 2001, pp. 260–79, for a thorough technical review of this area).

Hypnotic susceptibility has also been found to correlate with a range of reported paranormal and anomalous experiences (e.g., Atkinson, 1994; Nadon & Kihlstrom, 1987; Palmer & Van Der Velden, 1983; Pekala *et al.*, 1995; Spanos & Moretti, 1988; Wagner & Ratzeberg, 1987; Wickramasekera, 1989, pp. 19–35), although once again there are occasional studies that fail to find such a relationship (e.g., Persinger & De Sano, 1986). Richards (1990, p. 35) reported 'low and marginally significant' correlations between hypnotic susceptibility and self-reports of psychic experiences. Studies have also compared groups differing

in hypnotic susceptibility and have found differences in the degree to which anomalous/paranormal experiences are reported (e.g., Pekala *et al.*, 1992; 1995).

Absorption correlates moderately with paranormal belief (Palmer & Van Der Velden, 1983), subjective paranormal experiences (e.g., Irwin, 1981a) and mystical experiences (Spanos & Moretti, 1988). Both Irwin (1981b) and Myers *et al.* (1983) found that students who reported OBEs also demonstrated higher levels of absorption than those who did not. Irwin (1985) showed that the need for absorption was higher in experients than non-experients for a wide range of subjective paranormal experiences.

Dissociativity has often been shown to be correlated with paranormal belief (e.g., Greening, 2002, Study 2.2; Irwin, 1994; Pekala *et al.*, 1995; Wolfradt, 1997), but some studies have failed to find such a relationship (Greening, 2002, Study 2.1; Groth-Marnat *et al.*, 1998–99). Makasovski and Irwin (1999) present data suggesting that pathological dissociation predicts belief in parapsychological and spiritual concepts, but that non-pathological dissociative tendencies (absorption) do not correlate with paranormal belief. Rattet and Bursik (2001) reported that dissociative tendencies were related to paranormal belief, but not to self-reported precognitive experiences. Dissociativity has been shown to be related to the tendency to report a wide range of paranormal and anomalous experiences (Pekala *et al.*, 1995; Richards, 1991; Ross & Joshi, 1992; Ross *et al.*, 1991). Powers (1994) has shown that a group of alleged alien abductees showed higher levels of dissociativity than a matched sample of non-abductees. Children reporting past-life memories have been shown to have higher levels of dissociative tendencies in both Sri Lanka (Haraldsson *et al.*, 2000) and Lebanon (Haraldsson, 2002). Greyson (2000) has reported that although people reporting near-death experiences (NDEs) are psychologically healthy, some do show non-pathological signs of dissociation. The possibility that at least some reports of NDEs may be based upon false memories was raised by French (2001c) in a commentary on a prospective study of NDEs by van Lommel *et al.* (2001). The latter investigators interviewed a number of patients two years after they had suffered cardiac arrests that they had reported at the time were not associated with NDEs. At the two-year follow-up interviews, however, four of the 37 patients now reported that they had indeed experienced NDEs during their cardiac arrest.

In summary, it appears that there are numerous studies supporting an association between paranormal beliefs and reports of anomalous experiences on the one hand and a range of psychological factors thought to be associated with increased susceptibility to false memories on the other. It is important at this point to emphasise, however, that this pattern of correlations is also consistent with an alternative interpretation, one that is taken seriously by many parapsychologists. It is possible that individuals who score highly on such measures as dissociativity, hypnotic susceptibility and so on have the right psychological profile to experience genuine paranormal phenomena (if they actually exist). Of course, the false-memory hypothesis and the psi hypothesis are not mutually

exclusive. The correct interpretation of the pattern of findings will only be resolved by empirical investigation.

Having established a *prima facie* case for a link between false memories and paranormal beliefs and tendency to report anomalous experiences, we shall now review the relatively few studies that have investigated the postulated link directly. Haraldsson (1985) reported a low but significant correlation between suggestibility (as measured by the Gudjonsson Suggestibility Scale) and global paranormal beliefs (as measured by Tobacyk's Paranormal Belief Scale, PBS). Of the seven sub-scales of the PBS, only those measuring belief in witchcraft, spiritualism and precognition were significantly correlated with suggestibility.

Blackmore and Rose (1997) tested the hypothesis that susceptibility to false memories would be correlated with paranormal belief using a reality-monitoring task. Participants were initially shown simple drawings of objects or asked to imagine drawings of objects. Over subsequent sessions spanning a number of weeks, they were questioned regarding their memory of the pictures (both real and imagined). In a final session, they were asked to indicate whether each drawing had initially been presented or imagined. A false memory was recorded every time a picture that had only been imagined was recorded as having been presented. No correlation was found between susceptibility to false memories and paranormal belief. Three similar experiments by Rose and Blackmore (2001) also failed to find the predicted relationship. Greening (2002), using a similar methodology, did find a significant correlation in the predicted direction, but was unable to replicate the effect in two follow-up experiments.

Clancy *et al.* (2002) used the word list paradigm of Roediger and McDermott (1995) in a study comparing people with recovered memories of alien abduction, people who believed they had been abducted but without such memories, and people who denied having been abducted by aliens. The group with memories of abduction were shown to be more susceptible to false memories than the control participants. Furthermore, false recognition and recall were correlated with hypnotic susceptibility, depressive symptoms and schizotypic features.

V: Directions for Future Research

It is clear that direct attempts to prove a link between susceptibility to false memories and reports of anomalous experiences have so far met with only limited success. However, this may reflect the methods that have been used to date to test the hypothesis. It is unlikely that all of the different measures of susceptibility to false memory are measuring the same thing. It would therefore be advisable if future studies concentrated mainly upon those techniques that would appear to be most relevant to the possibility that memories for certain types of event may sometimes be false. Intuitively, naturally occurring distortions of autobiographical memory and susceptibility to implanted memories of entire episodes would seem to be the most relevant. Measurements of reality monitoring errors would appear to be of less relevance unless it could be shown that errors made in the task used by Blackmore and Rose (1997) generalise to more serious confusions

(such as between daydreams and reality). Susceptibility to misinformation is of some relevance, but one assumes that in everyday life it would be relatively rare for another individual to try to deliberately manipulate someone else's memory. Unintentional distortion by discussion with another individual is always a possibility, however. Finally, it is ironic that the word list paradigm is one of the few which seem to have been successful in discriminating between a group who had reported a particular anomalous experience and control groups (Clancy *et al.*, 2002), given the apparent lack of ecological validity of the task itself. This important finding awaits replication, however.

Another possible reason for the inconsistency in results to date is that many investigations have focussed upon belief in anomalous phenomena rather than reported experiences of anomalous phenomena. Although one of the most common reasons given for belief in the paranormal is personal experience, it is by no means the case that all believers have had such personal experience. There are many other reasons for belief in anomalous phenomena including media reports, personal accounts from trusted others, and so on. Clearly, one would expect a higher correlation between susceptibility to false memories and actual reports of particular anomalous experiences rather than belief in those anomalous phenomena. A further recommendation for future research in this area is that greater emphasis should be placed upon searching for correlates of the tendency to report anomalous phenomena as opposed to simply believing in them.

As is usually the case when considering psychological factors associated with paranormal and related beliefs, the studies reviewed above are generally quasi-experimental in nature. Participants cannot be randomly assigned to high and low paranormal belief groups. It is possible that susceptibility to false memories causes people to come to believe they have had a paranormal experience (even if they have not) which then produces or reinforces their belief in some particular aspect of the paranormal. On the other hand, it is reasonable to argue that pre-existing beliefs play a causal role in the acceptance of potential false memories as authentic. According to Hyman and Kleinknecht (1998), plausibility is an important factor in making such decisions. Whereas a fleeting memory of an ostensibly anomalous experience might be dismissed as probably being the memory of a dream by a sceptic, a believer is more likely to accept that it may reflect something that actually happened. Further reflection and elaboration may then lead to a more detailed and vivid 'memory'.

It must be emphasised, however, that memory distortion and the formation of false memories can never provide a complete explanation for all reports of anomalous events, nor is it intended to do so. A wide range of other factors needs to be taken into account (see, e.g., Cardeña *et al.*, 2000; French, 1992; Roberts & Groome, 2001; Zusne & Jones, 1989). To take but one example, a sincere report of having seen a ghost may well actually be a more or less accurate report of an hallucinatory experience. It is possible that the intense emotion generated by the experience may lead to less reliable testimony (e.g., Loftus, 1979) but the report is best understood by considering primarily the psychology of hallucinations (Bentall, 2000, pp. 85–120).

It is possible that much of what we take to be our personal autobiographical history is based upon false, or at least distorted, memories. This usually is not drawn to our attention because no one is likely to challenge mundane memories of ordinary everyday events unless one person's memory actually directly contradicts another. With respect to paranormal and related claims, however, the situation is entirely different. A listener may decide that a particular account must be inaccurate simply because the account contradicts that person's understanding of what is and what is not possible. Is it reasonable that such a person, without any claim whatsoever to first-hand knowledge of the events in question, should feel justified in adopting this sceptical position? On the basis of the evidence reviewed above, the answer has to be affirmative.

The review of recent developments in the area of false memory research suggests that a *prima facie* case can be made for a possible link between susceptibility to false memories and tendency to report anomalous experiences. One of the most important factors associated with paranormal and related beliefs is alleged personal experience of anomalous events and thus an indirect link may exist between susceptibility to false memories and level of belief. Alternatively, as described above, it may be that pre-existing beliefs play an important role in determining whether potential false memories are accepted as records of events that really occurred. They may also play a role in determining the content of such memories, as schema-driven distortions are likely to occur. To date, the few direct tests of the postulated links between susceptibility to false memories, reports of anomalous experiences, and level of paranormal and related beliefs have met with only limited success, but further research, taking into account the issues discussed above, is certainly warranted.

References

Appelle, S., Lynn, S.J., & Newman, L. (2000), 'Alien abduction experiences', in *Varieties of Anomalous Experience: Examining the scientific evidence*, ed. Cardeña, E., Lynn S.J., & Krippner, S. (Washington, DC: American Psychological Association).

Atkinson, R.P. (1994), 'Relationship of hypnotic susceptibility to paranormal beliefs and claimed experiences: Implications for hypnotic absorption', *American Journal of Clinical Hypnosis*, **37**, pp. 34–40.

Baker, R.A. (1992), *Hidden Memories: Voices and Visions from Within* (Buffalo, NY: Prometheus).

Barnier, A.J., & McConkey, K.M. (1992), 'Reports of real and false memories: The relevance of hypnosis, hypnotizability, and test control', *Journal of Abnormal Psychology*, **101**, pp. 521–7.

Bentall, R.P. (2000), 'Hallucinatory experiences', in *Varieties of Anomalous Experience: Examining the scientific evidence*, ed. Cardeña, E., Lynn, S.J., & Krippner, S. (Washington, DC: American Psychological Association).

Besterman, T. (1932), 'The psychology of testimony in relation to paraphysical phenomena: Report of an experiment', *Proceedings of the Society for Psychical Research*, **40**, pp. 363–87.

Blackmore, S.J. (1982), *Beyond the Body: An Investigation of Out-of-the-Body Experiences* (London: Heinemann).

Blackmore, S.J. (1986), 'Where am I? Perspectives in imagery, memory, and the OBE', in *Research in Parapsychology 1985*, ed. Weiner, D.H, & Radin, D.I. (Metuchen, NJ: Scarecrow Press).

Blackmore, S. (1994), 'Alien abduction: The inside story', *New Scientist*, No. 1952, 19 November, pp. 29–31.

Blackmore, S.J., & Rose, N. (1997), 'Reality and imagination: A psi-conducive confusion?' *Journal of Parapsychology*, **61**, pp. 321–35.

Brainerd, C.J., Reyna, V.F., & Poole, D.A. (2000), 'Fuzzy-trace theory and false memory: Memory theory in the courtroom', in *False-Memory Creation in Children and Adults: Theory, Research and Implications*, ed. Bjorklund, D.F. (Mahwah, NJ: Erlbaum).

Brown, R., & Kulik, J. (1977), 'Flashbulb memories', *Cognition*, **5**, pp. 73–99.

Cardeña, E., Lynn S.J., & Krippner, S. (eds.) (2000), *Varieties of Anomalous Experience: Examining the scientific evidence* (Washington, DC: American Psychological Association).

Clancy, S.A., McNally, R.J., Schacter, D.L., Lenzenweger, M.F., & Pitman, R.K. (2002), 'Memory distortion in people reporting abduction by aliens', *Journal of Abnormal Psychology*, **111**, pp. 455–61.

Cohen, G. (1989), *Memory in the Real World* (London: Erlbaum).

Conway, M.A. (ed.) (1997), *Recovered Memories and False Memories* (Oxford: Oxford University Press).

Cook, A.M., & Irwin, H.J. (1983), 'Visuospatial skills and the out-of-body experience', *Journal of Parapsychology*, **47**, pp. 23–35.

Davies, G., & Dalgleish, T. (eds.) (2001), *Recovered Memories: Seeking the Middle Ground* (Chichester: Wiley).

Deese, J. (1959), 'On the prediction of occurrence of particular verbal intrusions in immediate recall', *Journal of Experimental Psychology*, **58**, pp. 17–22.

Diamond, M.J., & Taft, R. (1975), 'The role played by ego-permissiveness and imagery in hypnotic responsivity', *International Journal of Clinical and Experimental Hypnosis*, **23**, pp. 130–8.

Dobson, M., & Markham, R. (1993), 'Imagery ability and source monitoring: Implications for eyewitness memory', *British Journal of Psychology*, **84**, pp. 111–18.

Eisen, M.L., & Carlson, E.B. (1998), 'Individual differences in suggestibility: Examining the influence of dissociation, absorption, and a history of childhood abuse', *Applied Cognitive Psychology*, **12**, pp. S47–61.

Femina, D.D., Yeager, C.A., & Lewis, D.O. (1990), 'Child abuse: Adolescent records vs. adult recall', *Child Abuse and Neglect*, **14**, pp. 227–31.

Finch, S.E. (2002), 'Daydream believers? Fantasy-proneness, transliminality and reality monitoring: A search for vulnerability factors in false memory creation', unpublished PhD thesis, Goldsmiths College, University of London.

French, C.C. (1992), 'Factors underlying belief in the paranormal: Do sheep and goats think differently?' *The Psychologist*, **5**, pp. 295–9.

French, C.C. (2001a), *Paranormal Perception? A Critical Evaluation* (London: Institute for Cultural Research. Monograph Series, No. 42).

French, C.C. (2001b), 'Alien abductions', in *Parapsychology: The Psychology of Unusual Experience*, ed. Roberts, R., & Groome, D. (London: Arnold).

French, C.C. (2001c), 'Dying to know the truth: Visions of a dying brain, or false memories?' *Lancet*, **358**, pp. 2010–11.

French, C.C., & Richards, A. (1993), 'Clock this! An everyday example of a schema-driven error in memory', *British Journal of Psychology*, **84**, pp. 249–53.

Gabbert, F., Memon, A., & Allan, K. (in press), 'Memory conformity: Can eyewitnesses influence each other's memories for an event?', *Applied Cognitive Psychology*.

Gabbert, F., Memon, A., & Allan, K. (submitted), 'Say it to my face: Examining the effects of socially encountered misinformation'.

Garry, M., Manning, C.G., Loftus, E.F., & Sherman, S.J. (1996), 'Imagination inflation: Imagining a childhood event inflates confidence that it occurred', *Psychonomic Bulletin & Review*, **3**, pp. 208–14.

Greening, E.K. (2002), 'The relationship between false memory and paranormal belief', Unpublished PhD thesis, University of Hertfordshire.

Grey, W. (1994), 'Philosophy and the paranormal. Part 2: Skepticism, miracles, and knowledge', *Skeptical Inquirer*, **18**, pp. 288–94.

Greyson, B. (2000), 'Dissociation in people who have near-death experiences: Out of their bodies or out of their minds?' *Lancet*, **355**, pp. 460–3.

Groth-Marnat, G., Roberts, L., & Ollier, K. (1998–99), 'Hypnotizability, dissociation, paranormal beliefs', *Imagination, Cognition and Personality*, **18**, pp. 127–32.

Haraldsson, E. (1985), 'Interrogative suggestibility and its relationship with personality, perceptual defensiveness and extraordinary beliefs', *Personality and Individual Differences*, **6**, pp. 765–7.

Haraldsson, E. (2002), 'Children who speak of past-life experiences: Is there a psychological explanation?' Paper presented to the Parapsychological Association Congress, Paris, August 2002.

Haraldsson, E., Fowler, P.C., & Periyannanpillai, V. (2000), 'Psychological characteristics of children who speak of a previous life: A further field study in Sri Lanka', *Transcultural Psychiatry*, **37**, pp. 525–44.

Harris, M. (1986), *Sorry, You've Been Duped! The Truth Behind Classic Mysteries of the Paranormal* (London: Weidenfeld and Nicolson).

Heaps, C., & Nash, M. (1999), 'Individual differences in imagination inflation', *Psychonomic Bulletin & Review*, **6**, pp. 313–18.

Holden, K.J., & French, C.C. (2002), 'Alien abduction experiences: Clues from neuropsychology and neuropsychiatry', *Cognitive Neuropsychiatry*, **7**, pp. 163–78.

Holmes, D.S. (1990), 'The evidence for repression: An examination of sixty years of research', in *Repression and dissociation: Implications for personality theory, psychopathology, and health*, ed. Singer, J.L. (Chicago and London: University of Chicago Press).

Hopkins, B., Jacobs, D.M., & Westrum, R. (1992), *Unusual personal experiences: An analysis of the the data from three national surveys conducted by the Roper Organisation* (Las Vegas, NV: Bigelow Holding Corporation).

Horselenberg, R., Merckelbach, H., Muris, P., Rassin, E., Sijsenaar, M., & Spaan, V. (2000), 'Imagining fictitious childhood events: The role of individual differences in imagination inflation', *Clinical Psychology and Psychotherapy*, **7**, pp. 128–37.

Hyman, I.E., Jr., & Billings, F.J. (1998), 'Individual differences and the creation of false childhood memories', *Memory*, **6**, pp. 1–20.

Hyman, I.E., Jr., Husband, T.H., & Billings, F.J. (1995), 'False memories of childhood experiences', *Applied Cognitive Psychology*, **9**, pp. 181–97.

Hyman, I.E., Jr., & Kleinknecht, E.E. (1999). 'False childhood memories: Research, theory, and applications', in *Trauma and Memory*, ed. Williams, L.M., & Banyard, V.L. (Thousand Oaks, CA: Sage).

Hyman, R. (1985), 'A critical historical overview of parapsychology', in *A Skeptic's Handbook of Parapsychology*, ed. Kurtz, P. (Buffalo, NY: Prometheus).

Irwin, H.J. (1981a), 'The psychological function of out-of-body experience: So who needs the out-of-body experience?' *Journal of Nervous and Mental Disease*, **169**, pp. 244–8.

Irwin, H. J. (1981b), 'Some psychological dimensions of the out-of-body experience', *Parapsychology Review*, **12**(4), pp. 1–6.

Irwin, H.J. (1985), 'Parapsychological phenomena and the absorption domain', *Journal of the American Society for Psychical Research*, **79**, pp. 1–11.

Irwin, H.J. (1994), 'Paranormal belief and proneness to dissociation', *Psychological Reports*, **75**, pp. 1344–6.

Iverson, J. (1977), *More Lives than One?* (London: Pan Books).

Jacobs, D.M. (1992), *Secret Life: Firsthand Accounts of UFO Abductions* (New York: Simon & Schuster).

Jacobs, D.M., & Hopkins, B. (1992). 'Suggested Techniques for Hypnosis and Therapy of Abductees', Unpublished manuscript, Temple University, Philadelphia.

Johnson, M.K., Hashtroudi, S., & Lindsay, D.S. (1993), 'Source monitoring', *Psychological Bulletin*, **114**, pp. 3–28.

Johnson, M.K., & Raye, C. (1981), 'Reality monitoring', *Psychological Review*, **88**, pp. 67–85.

Jones, W.H., & Russell, D. (1980), 'The selective processing of belief disconfirming information', *European Journal of Social Psychology*, **10**, pp. 309–12.

Kebbell, M.R., & Wagstaff, G.F. (1998), 'Hypnotic interviewing: The best way to interview eyewitnesses?', *Behavioral Sciences and the Law*, **16**, pp. 115–29.

Kumar, V. K., & Pekala, R.J. (2001), 'Relation of hypnosis-specific attitudes and behaviors to paranormal beliefs and experiences: A technical review', in *Hauntings and Poltergeists: Multidisciplinary Perspectives*, ed. Houran, J. & Lange, R. (Jefferson, NC: McFarland & Co).

Laurence, J.R., & Perry, C. (1983), 'Hypnotically created memory among highly hypnotizable subjects', *Science*, **222**, pp. 523–4.

Lindsay, D. S., & Read, J.D. (1995), '"Memory work" and recovered memories of childhood sexual abuse: Scientific evidence and public, professional, and personal issues', *Psychology, Public Policy, and the Law*, **1**, pp. 846–908.

Loftus, E.F. (1975), 'Leading questions and the eyewitness report', *Cognitive Psychology*, **7**, pp. 560–72.

Loftus, E.F. (1979), *Eyewitness Testimony* (Cambridge, MA: Harvard University Press).

Loftus, E.F. (1993), 'The reality of repressed memories', *American Psychologist*, **48**, pp. 518–37.

Loftus, E.F. (2001), 'Imagining the past', *The Psychologist*, **14**, pp. 584–7.

Loftus, E.F., Garry, M. & Feldman, J. (1994), 'Forgetting sexual trauma: What does it mean when 38% forget?' *Journal of Consulting and Clinical Psychology*, **62**, pp. 1177–81.

Loftus, E.F., & Ketcham, K. (1994), *The Myth of Repressed Memory: False Memories and Allegations of Sexual Abuse* (New York: St. Martin's Press).

Loftus, E.F., Miller, D.G., & Burns, H.J. (1978), 'Semantic integration of verbal information into a visual memory', *Journal of Experimental Psychology: Human Learning and Memory*, **4**, pp. 19–31.

Loftus, E.F., & Pickrell, J.E. (1995), 'The formation of false memories', *Psychiatric Annals*, **25**, pp. 720–5.

Loftus, E.F., Polonsky, S., & Fullilove, M.T. (1994), 'Memories of childhood sexual abuse: Remembering and repressing', *Psychology of Women Quarterly*, **18**, pp. 67–84.

Lynn, S.J., & McConkey, K.M. (eds.) (1998), *Truth in Memory* (New York & London: Guilford Press).

Makasovski, T., & Irwin, H.J. (1999), 'Paranormal belief, dissociative tendencies, and parental encouragement of imagination in childhood', *Journal of the American Society for Psychical Research*, **93**, pp. 233–47.

Markham, R., & Hynes, L. (1993), 'The effect of vividness of imagery on reality monitoring', *Journal of Mental Imagery*, **17**, pp. 159–70.

Mazzoni, G.A.L., & Loftus, E.F. (1998), 'Dream interpretation can change beliefs about the past', *Psychotherapy*, **35**, pp. 177–87.

McCloskey, M., & Zaragoza, M.S. (1985). 'Misleading post-event information and memory for events: Arguments and evidence against memory impairment hypotheses', *Journal of Experimental Psychology: General*, **114**, pp. 1–16.

McConkey, K.M., Barnier, A.J., & Sheehan, P.W. (1998), 'Hypnosis and pseudomemory: Understanding the findings and their implications', in *Truth in Memory,* ed. Lynn, S.J., & McConkey, K.M. (New York & London: Guilford Press).

Myers, S.A., Austrin, H.R., Grisso, J.T., & Nickeson, R.C. (1983), 'Personality characteristics as related to the out-of-body experience', *Journal of Parapsychology,* **47**, pp. 131–44.

Nadon, R., & Kihlstrom, J.F. (1987), 'Hypnosis, psi, and the psychology of anomalous experience', *Behavioural and Brain Sciences,* **10**, pp. 597–9.

Nadon, R., Laurence, J.R., & Perry, C. (1987), 'Multiple predictors of hypnotic susceptibility', *Journal of Personality and Social Psychology,* **53**, pp. 948–60.

Neisser, U., & Harsch, N. (1993), 'Phantom flashbulbs: False recollections of hearing the news about Challenger', in *Affect and Accuracy in Recall: Studies of 'flashbulb memories'*, ed. Winograd, E., & Neisser, U. (New York: Cambridge University Press).

Ofshe, R., & Watters, E. (1994), *Making Monsters: False Memories, Psychotherapy, and Sexual Hysteria* (New York: Scribner).

Orne, M.T. (1979), 'The use and misuse of hypnosis in court', *International Journal of Clinical and Experimental Hypnosis,* **27**, pp. 311–41.

Palmer, J., & Van Der Velden, I. (1983), 'ESP and "hypnotic imagination": A group free-response study', *European Journal of Parapsychology,* **4**, pp. 413–34.

Pekala, R.J., Kumar, V.K., & Cummings, J. (1992), 'Types of high hypnotically susceptible individuals and reported attitudes and experiences of the paranormal and the anomalous', *Journal of the American Society for Psychical Research,* **86**, pp. 135–50.

Pekala, R.J., Kumar, V.K., & Marcano, G. (1995), 'Anomalous/paranormal experiences, hypnotic susceptibility, and dissociation', *Journal of the American Society for Psychical Research,* **89**, pp. 313–32.

Persinger, M.A., & De Sano, C.F. (1986), 'Temporal lobe signs: positive correlations with imaginings and hypnosis induction profiles', *Psychological Reports,* **58**, pp. 347–50.

Platt, R.D., Lacey, S.C., Iobst, A.D., & Finkelman, D. (1998), 'Absorption, dissociation, fantasy-proneness as predictors of memory distortion in autobiographical and laboratory-generated memories', *Applied Cognitive Psychology,* **12**, pp. S77–89.

Pope, K.S., & Brown, L.S. (1996), *Recovered Memories of Abuse: Assessment, Therapy, Forensics* (Washington, DC: American Psychological Association).

Powers, S.M. (1994), 'Dissociation in alleged extraterrestrial abductees', *Dissociation,* 7, pp. 44–50.

Pynoos, R.S., & Nader, K. (1989), 'Children's memory and proximity to violence', *Journal of the American Academy of Child and Adolescent Psychiatry,* **28**, pp. 236–41.

Rattet, S.L., & Bursik, K. (2001), 'Investigating the personality correlates of paranormal belief and precognitive experience', *Personality and Individual Differences,* **31**, pp. 433–44.

Richards, D.G. (1990), 'Hypnotic susceptibility and subjective psychic experiences: A study of participants in A.R.E. conferences', *Journal of Parapsychology,* **54**, pp. 35–51.

Richards, D.G. (1991), 'A study of the correlations between subjective psychic experiences and dissociative experiences', *Dissociation,* **4**, pp. 83–91.

Roberts, R., & Groome, D. (eds.) (2001), *Parapsychology: The Psychology of Unusual Experience* (London: Arnold).

Roediger, H.L, III, & McDermott, K.B. (1995), 'Creating false memories: Remembering words not presented on lists', *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **21**, pp. 803–14.

Rose, N., & Blackmore, S.J. (2001), 'Are false memories psi-conducive?' *Journal of Parapsychology,* **65**, pp. 125–44.

Ross, C.A., & Joshi, S. (1992), 'Paranormal experiences in the general population', *Journal of Nervous and Mental Disease,* **180**, pp. 357–61.

Ross, C.A., Ryan, L., Voigt, H., & Eide, L. (1991), 'High and low dissociators in a college student population', *Dissociation,* **4**, pp. 147–51.

Schacter, D.L. (ed.) (1995), *Memory Distortion: How Minds, Brains, and Societies Reconstruct the Past* (Cambridge, MA & London: Harvard University Press).

Sheehan, P.W., Statham, D., & Jamieson, G.A. (1991), 'Pseudomemory effects and their relationship to level of susceptibility to hypnosis and state instruction', *Journal of Personality and Social Psychology,* **60**, pp. 130–7.

Spanos, N.P. (1996), *Multiple Identities and False Memories: A Sociocognitive Perspective* (Washington, DC: American Psychological Association).

Spanos, N.P., Burgess, C.A. & Burgess, M.F. (1994), 'Past-life identities, UFO abductions, and Satanic ritual abuse: The social construction of memories', *International Journal of Clinical and Experimental Hypnosis,* **42**, pp. 433–46.

Spanos, N.P, Menary, E., Gabora, N.J., DuBreuil, S.C. & Dewhirst, B. (1991), 'Secondary identity enactments during hypnotic past-life regression: A sociocognitive perspectiv", *Journal of Personality and Social Psychology,* **61**, pp. 308–20.

Spanos, N.P., & Moretti, P. (1988), 'Correlates of mystical and diabolical experiences in a sample of female university students', *Journal for the Scientific Study of Religion,* **27**, pp. 105–16.

Tellegen, A., & Atkinson, G. (1974), 'Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility, *Journal of Abnormal Psychology*, **83**, pp. 268–77.
Tomes, J.L., & Katz, A.N. (1997), 'Habitual susceptibility to misinformation and individual differences in eyewitness memory', *Applied Cognitive Psychology*, **11**, pp. 233–51.
Van Lommel, P., van Wees, R., Meyers, V., & Elfferich, I. (2001), 'Near-death experience in survivors of cardiac arrest: A prospective study in the Netherlands', *Lancet*, **358**, pp. 2039–45.
Wagner, M.W., & Ratzeberg, F.H. (1987), 'Hypnotic suggestibility and paranormal belief', *Psychological Reports*, **60**, pp. 1069–70.
Wagstaff, G.F. (1989), 'Forensic aspects of hypnosis', in *Hypnosis: The cognitive–behavioral perspective*, ed. Spanos, N.P., & Chaves, J.F. (Buffalo, NY: Prometheus).
Wickramasekera, I. (1989), 'Risk factors for parapsychological verbal reports, hypnotizability and somatic complaints', in *Parapsychology and Human Nature*, ed. Shapin, B., & Coly, L. (New York: Parapsychology Foundation).
Williams, L.M. (1994), 'Recall of childhood trauma: A prospective study of women's memories of child sexual abuse', *Journal of Consulting and Clinical Psychology*, **62**, pp. 1167–76.
Winograd, E., Peluso, J.P., & Glover, T.A. (1998), 'Individual differences in susceptibility to memory illusions', *Applied Cognitive Psychology*, **12**, pp. S5–27.
Wiseman, R., & Morris, R.L. (1995), 'Recalling pseudo-psychic demonstrations', *British Journal of Psychology*, **86**, pp.113–25.
Wiseman, R., Smith, M., & Wiseman, J. (1995), 'Eyewitness testimony and the paranormal', *Skeptical Inquirer*, **19**(6), pp. 29–32.
Wolfradt, U. (1997), 'Dissociative experiences, trait anxiety and paranormal beliefs', *Personality and Individual Differences*, **23**, pp. 15–19.
Yapko, M.D. (1994), *Suggestions of Abuse* (New York: Simon & Schuster).
Zusne, L., & Jones, W.H. (1989), *Anomalistic Psychology: A Study of Magical Thinking* (Hillsdale, NJ: Erlbaum).

[4]

Birthmarks and Birth Defects Corresponding to Wounds on Deceased Persons

IAN STEVENSON

Department of Psychiatric Medicine, University of Virginia, School of Medicine, Charlottesville, Virginia 22908

Abstract — Almost nothing is known about why pigmented birthmarks (moles or nevi) occur in particular locations of the skin. The causes of most birth defects are also unknown. About 35% of children who claim to remember previous lives have birthmarks and/or birth defects that they (or adult informants) attribute to wounds on a person whose life the child remembers. The cases of 210 such children have been investigated. The birthmarks were usually areas of hairless, puckered skin; some were areas of little or no pigmentation (hypopigmented macules); others were areas of increased pigmentation (hyperpigmented nevi). The birth defects were nearly always of rare types. In cases in which a deceased person was identified the details of whose life unmistakably matched the child's statements, a close correspondence was nearly always found between the birthmarks and/or birth defects on the child and the wounds on the deceased person. In 43 of 49 cases in which a medical document (usually a postmortem report) was obtained, it confirmed the correspondence between wounds and birthmarks (or birth defects). There is little evidence that parents and other informants imposed a false identity on the child in order to explain the child's birthmark or birth defect. Some paranormal process seems required to account for at least some of the details of these cases, including the birthmarks and birth defects.

Key words: Birthmarks, birth defects, paranormal processes, teratology

Introduction

Although counts of moles (hyperpigmented nevi) have shown that the average adult has between 15 and 18 of them (Pack and Davis, 1956), little is known about their cause — except for those associated with the genetic disease neurofibromatosis — and even less is known about why birthmarks occur in one location of the body instead of in another. In a few instances a genetic factor has been plausibly suggested for the location of nevi (Cockayne, 1933; Denaro, 1944; Maruri, 1961); but the cause of the location of most birthmarks remains unknown. The causes of many, perhaps most, birth defects remain similarly unknown. In large series of birth defects in which investigators have searched for the known causes, such as chemical teratogens (like thalidomide), viral infections, and genetic factors, between 43% (Nelson and

Presented at the Eleventh Annual Meeting of the Society for Scientific Exploration held at Princeton University, June 11-13, 1992.

Holmes, 1989) and 65 — 70% (Wilson, 1973) of cases have finally been assigned to the category of "unknown causes."

Among 895 cases of children who claimed to remember a previous life (or were thought by adults to have had a previous life), birthmarks and/or birth defects attributed to the previous life were reported in 309 (35%) of the subjects. The birthmark or birth defect of the child was said to correspond to a wound (usually fatal) or other mark on the deceased person whose life the child said it remembered. This paper reports an inquiry into the validity of such claims. With my associates I have now carried the investigation of 210 such cases to a stage where I can report their details in a forthcoming book (Stevenson, forthcoming). This article summarizes our findings.

Children who claim to remember previous lives have been found in every part of the world where they have been looked for (Stevenson, 1983; 1987), but they are found most easily in the countries of South Asia. Typically, such a child begins to speak about a previous life almost as soon as it can speak, usually between the ages of two and three; and typically it stops doing so between the ages of five and seven (Cook, Pasricha, Samararatne, Win Maung, and Stevenson, 1983). Although some of the children make only vague statements, others give details of names and events that permit identifying a person whose life and death corresponds to the child's statements. In some instances the person identified is already known to the child's family, but in many cases this is not so. In addition to making verifiable statements about a deceased person, many of the children show behavior (such as a phobia) that is unusual in their family but found to correspond to behavior shown by the deceased person concerned or conjecturable for him (Stevenson, 1987; 1990).

Although some of the birthmarks occurring on these children are "ordinary" hyperpigmented nevi (moles) of which every adult has some (Pack and Davis, 1956), most are not. Instead, they are more likely to be puckered and scarlike, sometimes depressed a little below the surrounding skin, areas of hairlessness, areas of markedly diminished pigmentation (hypopigmented macules), or port-wine stains *(nevi flammei)*. When a relevant birthmark is a hyperpigmented nevus, it is nearly always larger in area than the "ordinary" hyperpigmented nevus. Similarly, the birth defects in these cases are of unusual types and rarely correspond to any of the "recognizable patterns of human malformation" (Smith, 1982).

Methods

My investigations of these cases included interviews, often repeated, with the subject and with several or many other informants for both families. With rare exceptions, only firsthand informants were interviewed. All pertinent written records that existed, particularly death certificates and postmortem reports, were sought and examined. In the cases in which the informants said that the two families had no previous acquaintance, I made every effort to exclude all possibility that some information might nevertheless have passed normally to the child, perhaps through a half-forgotten mutual acquaintance of the two

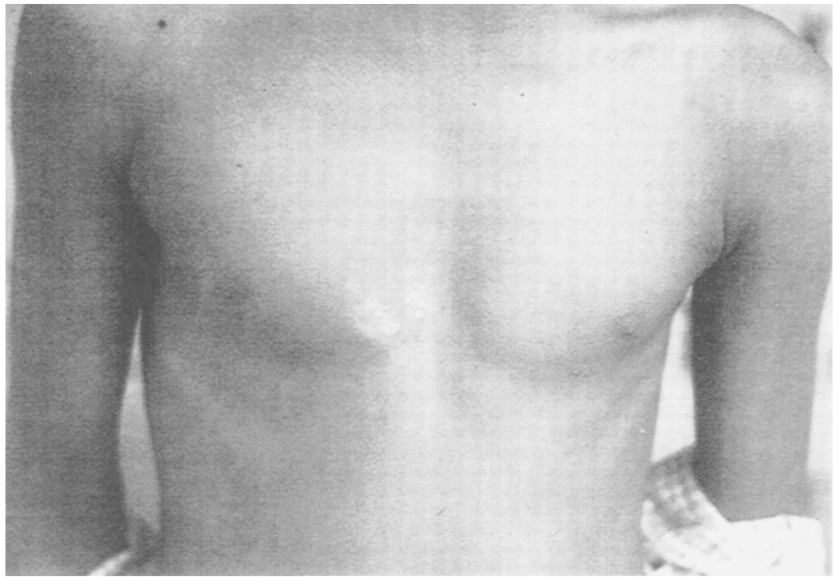

Fig. 1. Hypopigmented macule on chest of an Indian youth who, as a child, said he remembered the life of a man, Maha Ram, who was killed with a shotgun fired at close range.

families. I have published elsewhere full details about methods (Stevenson, 1975; 1987).

I did not accept any indicated mark as a birthmark unless a firsthand witness assured me that it had been noticed immediately after the child's birth or, at most, within a few weeks. I enquired about the occurrence of similar birthmarks in other members of the family; in nearly every instance this was denied, but in seven cases a genetic factor could not be excluded.

Birth defects of the kind in question here would be noticed immediately after the child's birth. Inquiries in these cases excluded (again with rare exceptions) the known causes of birth defects, such as close biological relationship of the parents (consanguinity), viral infections in the subject's mother during her pregnancy, and chemical causes of birth defects like alcohol.

Results

Correspondences between Wounds and Birthmarks

A correspondence between birthmark and wound was judged satisfactory if the birthmark and wound were both within an area of 10 square centimeters at the same anatomical location; in fact, many of the birthmarks and wounds were much closer to the same location than this. A medical document, usually a postmortem report, was obtained in 49 cases. The correspondence between wound and birthmark was judged satisfactory or better by the mentioned criterion in 43 (88%) of these cases and not satisfactory in 6 cases. Several different

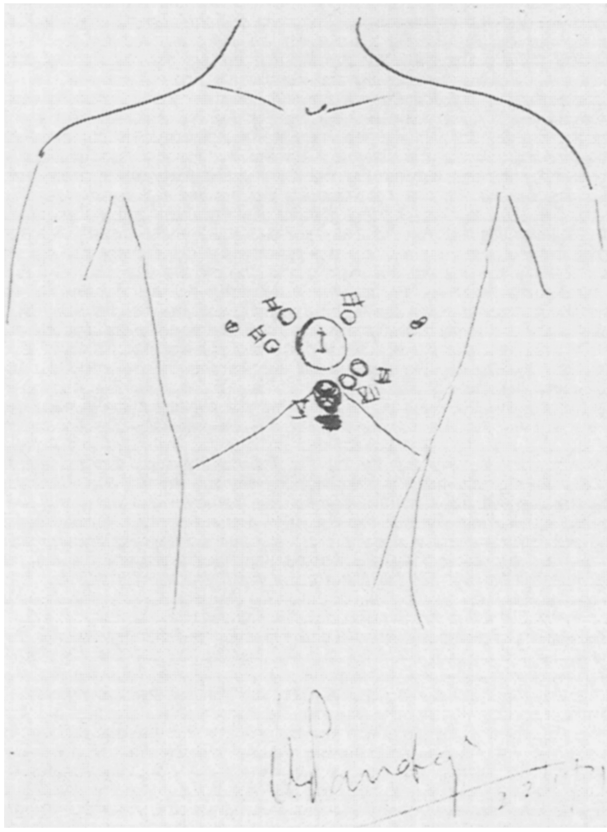

Fig. 2. The circles show the principal shotgun wounds on Maha Ram, for comparison with Figure 1.

explanations seem to be required to account for the discrepant cases, and I discuss these elsewhere (Stevenson, forthcoming). Figure 1 shows a birthmark (an area of hypopigmentation) on an Indian child who said he remembered the life of a man who had been killed with a shotgun fired at close range. Figure 2 shows the location of the wounds recorded by the pathologist. (The circles were drawn by an Indian physician who studied the postmortem report with me.)

The high proportion (88%) of concordance between wounds and birthmarks in the cases for which we obtained postmortem reports (or other confirming documents) increases confidence in the accuracy of informants' memories concerning the wounds on the deceased person in those more numerous cases for which we could obtain no medical document. Not all errors of informants' memories would have resulted in attributing a correspondence between birthmarks and wounds that did not exist; in four cases (possibly five) reliance on

Fig. 3. Large verrucous epidermal nevus on head of a Thai man who as a child said he remembered the life of his paternal uncle, who was killed with a blow on the head from a heavy knife.

an informant's memory would have resulted in missing a correspondence to which a medical document attested.

Cases with Two or More Birthmarks

The argument of chance as accounting for the correspondence between birthmarks and wounds becomes much reduced when the child has two or more birthmarks each corresponding to a wound on the deceased person whose life he claims to remember. Figure 3 shows a major abnormality of the skin (verrucous epidermal nevus) on the back of the head of a Thai man who, as a child, recalled the life of his uncle, who had been struck on the head with a heavy knife and killed almost instantly. The subject also had a deformed toe-

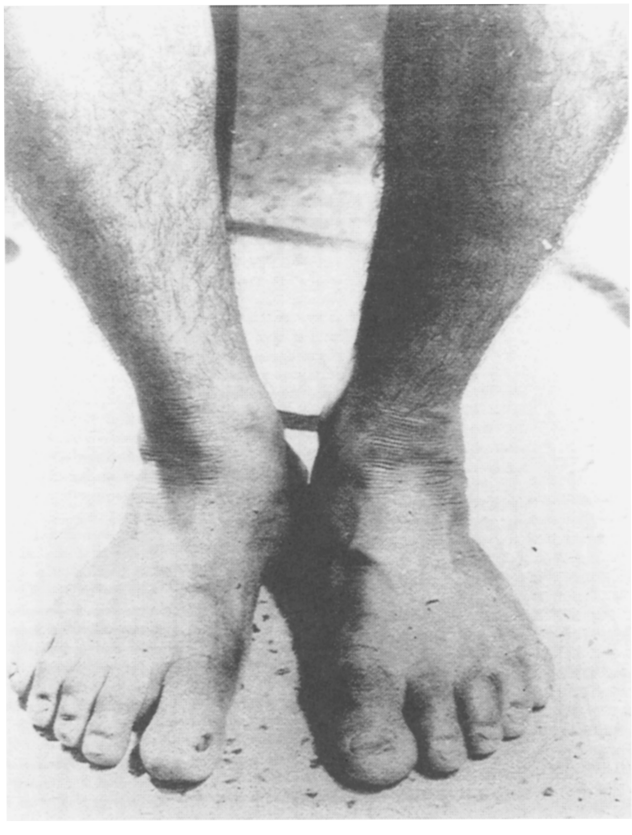

Fig. 4. Congenital malformation of nail on right great toe of the Thai subject shown in Figure 3. This malformation corresponded to a chronic ulcer of the right great toe from which the subject's uncle had suffered.

nail of the right great toe (Figure 4). This corresponded to a chronic infection of the same toe from which the subject's uncle had suffered for some years before he died.

The series includes 18 cases in which two birthmarks on a subject corresponded to gunshot wounds of entry and exit. In 14 of these one birthmark was larger than the other, and in 9 of these 14 the evidence clearly showed that the smaller birthmark (usually round) corresponded to the wound of entry and the larger one (usually irregular in shape) corresponded to the wound of exit. These observations accord with the fact that bullet wounds of exit are nearly always larger than wounds of entry (Fatteh, 1976; Gordon and Shapiro, 1982). Figure 5 shows a small round birthmark on the back of the head of a Thai boy, and Figure 6 shows a larger, irregularly shaped birthmark at the front of his head. The boy said that he remembered the life of a man who was shot in the head from behind. (The mode of death was verified, but no medical document was obtainable.) In addition to the 9 cases I have investigated myself, Mills re-

Fig. 5. Small, round puckered birthmark on a Thai boy that corresponded to the bullet wound of entry in a man whose life he said he remembered and who had been shot with a rifle from behind.

Fig. 6. Larger, irregularly shaped birthmark on the frontal area of the head of the Thai boy shown in Figure 5. This birthmark corresponded to the bullet wound of exit on the Thai man whose life the boy said he remembered.

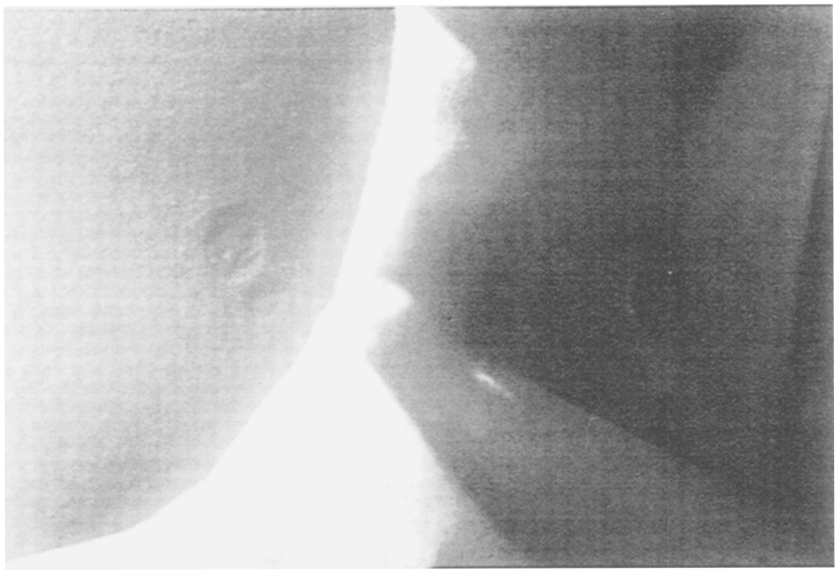

Fig. 7. Two round, puckered, scarlike birthmarks of different sizes on the left breast of a Burmese woman who as a child said she remembered the life of a woman who was fatally wounded by a shotgun that used a cartridge containing shot of different sizes.

ported another case having the feature of a small round birthmark (corresponding to the wound of entry) and a larger birthmark corresponding to the wound of exit (both verified by a postmortem report) (Mills, 1989).

I have calculated the odds against chance of two birthmarks correctly corresponding to two wounds. The surface area of the skin of the average adult male is 1.6 meters (Spalteholz, 1943). If we were to imagine this area square and spread on a flat surface, its dimensions would be approximately 127 centimeters by 127 centimeters. Into this area would fit approximately 160 squares of the size 10 centimeters square that I mentioned above. The probability that a single birthmark on a person would correspond in location to a wound within the area of any of the 160 smaller squares is only 1/160. However, the probability of correspondences between two birthmarks and two wounds would be $(1/160)^2$ i.e. 1 in 25,600. (This calculation assumes that birthmarks are uniformly distributed over all regions of the skin. This is incorrect [Pack, Lenson, and Gerber, 1952], but I believe the variation can be ignored for the present purpose.)

Examples of Other Correspondences of Detail between Wounds and Birthmarks

A Thai woman had three separate linear hypopigmented scarlike birthmarks near the midline of her back; as a child she had remembered the life of a woman who was killed when struck three times in the back with an ax. (Informants verified this mode of death, but no medical record was obtainable.) A woman of Burma was born with two perfectly round birthmarks in her left

chest (Figure 7); they slightly overlapped, and one was about half the size of the other. As a child she said that she remembered the life of a woman who was accidentally shot and killed with a shotgun. A responsible informant said the shotgun cartridge had contained shot of two different sizes. (No medical record was obtainable in this case.) Another Burmese child said that she remembered the life of her deceased aunt, who had died during surgery for congenital heart disease. This child had a long, vertical linear hypopigmented birthmark close to the midline of her lower chest and upper abdomen; this birthmark corresponded to the surgical incision for the repair of the aunt's heart. (I obtained a medical record in this case.) In contrast, a child of Turkey had a horizontal linear birthmark across the right upper quadrant of his abdomen. It resembled the scar of a surgeon's transverse abdominal incision. The child said that he remembered the life of his paternal grandfather, who had become jaundiced and was operated on before he died. He may have had a cancer of the head of the pancreas, but I could not learn a precise medical diagnosis. Two Burmese subjects remembered as children the lives of persons who had died after being bitten by venomous snakes, and the birthmarks of each corresponded to therapeutic incisions made at the sites of the snakebites on the persons whose lives they remembered. Another Burmese subject also said as a child that she remembered the life of a child who had been bitten on the foot by a snake and died. In this case, however, the child's uncle had applied a burning cheroot to the site of the bite — a folk remedy for snakebite in parts of Burma; and the subject's birthmark was round and located at the site on the foot where the bitten child's uncle had applied the cheroot.

Three Examples of Birth Defects

Figure 8 shows the right side of the head of a Turkish boy with a diminished and malformed ear (unilateral microtia). He also had underdevelopment of the right side of his face (hemifacial microsomia). He said that he remembered the life of a man who had been shot (with a shotgun) at point-blank range. The wounded man was taken to a hospital where he died 6 days later — of injuries to the brain caused by shot that had penetrated the right side of the skull. (I obtained a copy of the hospital record.)

Figure 9 shows fingers almost absent congenitally on one hand (unilateral brachydactyly) in a child of India who said he remembered the life of another child who had put his right hand into the blades of a fodder-chopping machine and lost his fingers. Most cases of brachydactyly involve only a shortening of the middle phalanges. In the present case there were no phalangeal bones, and the fingers were represented by mere stubs. Unilateral brachydactyly is exceedingly rare, and I have not found a published report of a case, although a colleague (plastic surgeon) has shown me a photograph of one case that came under his care.

Figure 10 shows congenital absence of the lower right leg (unilateral hemimelia) in a Burmese girl. She said that she remembered the life of a girl

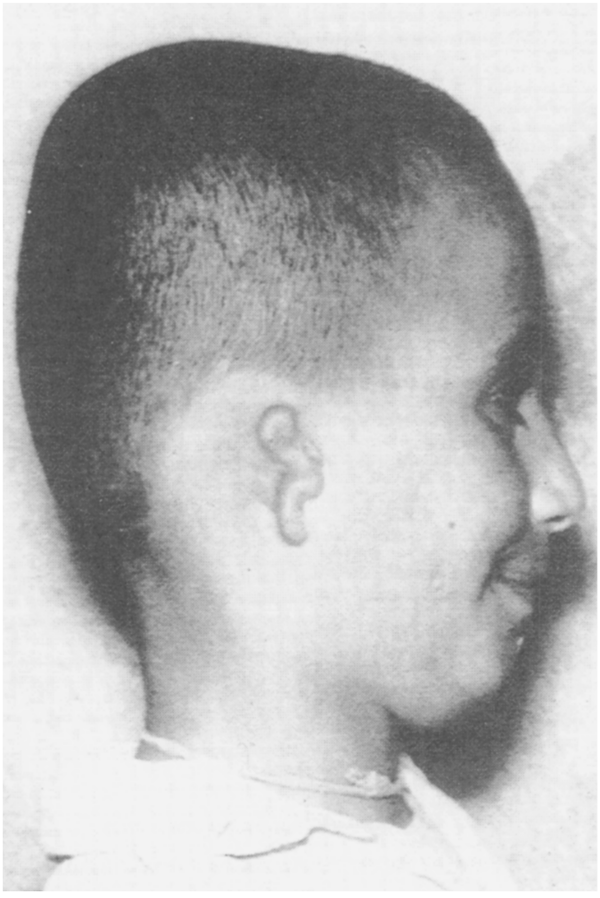

Fig. 8. Severely malformed ear (microtia) in a Turkish boy who said that he remembered the life of a man who was fatally wounded on the right side of the head by a shotgun discharged at close range.

who was run over by a train. Eyewitnesses said that the train severed the girl's right leg first, before running over the trunk. Lower hemimelia is an extremely rare condition, and Frantz and O'Rahilly (1961) found it in only 12 (4.0%) of 300 cases of all congenital skeletal deficiencies that they examined.

Discussion

Because most (but not all) of these cases develop among persons who believe in reincarnation, we should expect that the informants for the cases would interpret them as examples according with their belief; and they usually do. It is necessary, however, for scientists to think of alternative explanations.

The most obvious explanation of these cases attributes the birthmark or birth defect on the child to chance, and the reports of the child's statements

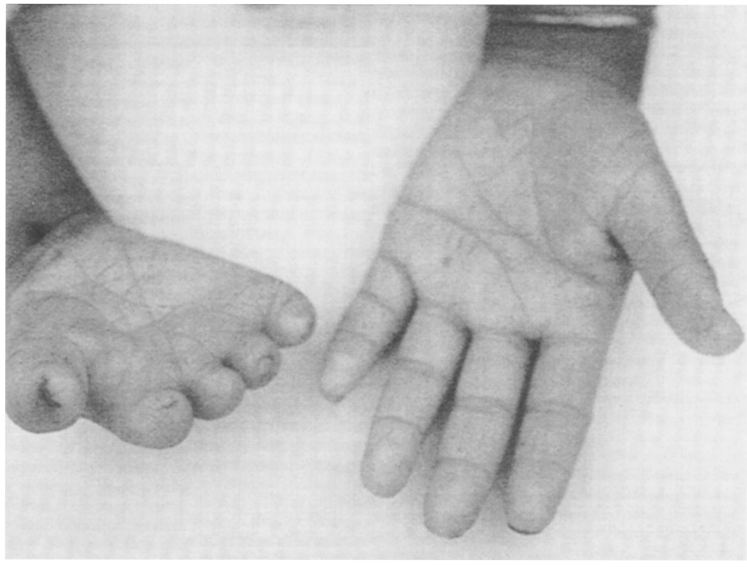

Fig. 9. Almost absent fingers (brachydactyly) of one hand in a boy of India who said he remembered the life of a boy of another village who had put his hand into the blades of a fodder-chopping machine and had its fingers amputated.

and unusual behavior then become a parental fiction intended to account for the birthmark (or birth defect) in terms of the culturally accepted belief in reincarnation. There are, however, important objections to this explanation. First, the parents (and other adults concerned in a case) have no need to invent and narrate details of a previous life in order to explain their child's lesion. Believing in reincarnation, as most of them do, they are nearly always content to attribute the lesion to *some* event of a previous life without searching for a *particular* life with matching details. Second, the lives of the deceased persons figuring in the cases were of uneven quality both as to social status and commendable conduct. A few of them provided models of heroism or some other enviable quality; but many of them lived in poverty or were otherwise unexemplary. Few parents would impose an identification with such persons on their children. Third, although in most cases the two families concerned were acquainted (or even related), I am confident that in at least 13 cases (among 210 carefully examined with regard to this matter) the two families concerned had never even heard about each other before the case developed. The subject's family in these cases can have had no information with which to build up an imaginary previous life which, it later turned out, closely matched a real one. In another 12 cases the child's parents had heard about the death of the person concerned, but had no knowledge of the wounds on that person. Limitations of space for this article oblige me to ask readers to accept my appraisal of these 25 cases for this matter; but in my forthcoming work I give a list of the

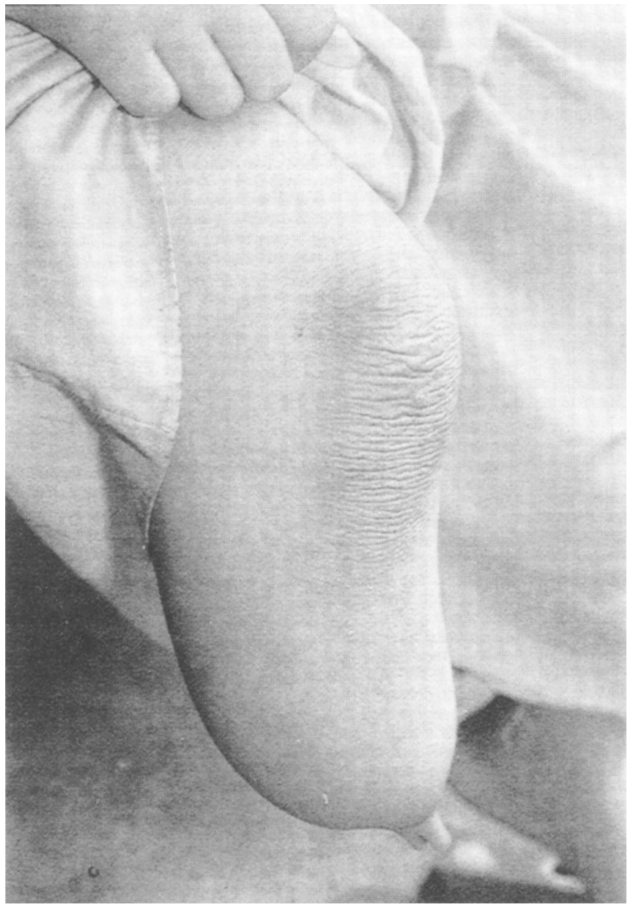

Fig. 10. Congenital absence of lower leg (unilateral hemimelia) in a girl of Burma who said she remembered the life of a young woman who was accidentally run over by a train, with her right leg being severed first.

cases from which readers can find the detailed reports of the cases and from reading them judge this important question for themselves. Fourth, I think I have shown that chance is an improbable interpretation for the correspondences in location between two or more birthmarks on the subject of a case and wounds on a deceased person.

Persons who reject the explanation of chance combined with a secondarily confected history may consider other interpretations that include paranormal processes, but fall short of proposing a life after death. One of these supposes that the birthmark or birth defect occurs by chance and the subject then by telepathy learns about a deceased person who had a similar lesion and devel-

ops an identification with that person. The children subjects of these cases, however, never show paranormal powers of the magnitude required to explain the apparent memories in contexts outside of their seeming memories.

Another explanation, which would leave less to chance in the production of the child's lesion, attributes it to a maternal impression on the part of the child's mother. According to this idea, a pregnant woman, having a knowledge of the deceased person's wounds, might influence a gestating embryo and fetus so that its form corresponded to the wounds on the deceased person. The idea of maternal impressions, popular in preceding centuries and up to the first decades of this one, has fallen into disrepute. Until my own recent article (Stevenson, 1992) there had been no review of series of cases since 1890 (Dabney, 1890); and cases are rarely published now (Williams and Pembroke, 1988). Nevertheless, some of the published cases — old and new — show a remarkable correspondence between an unusual stimulus in the mind of a pregnant woman and an unusual birthmark or birth defect in her later-born child. Also, in an analysis of 113 published cases I found that the stimulus occurred to the mother in the first trimester in 80 cases (Stevenson, 1992). The first trimester is well known to be the one of greatest sensitivity of the embryo/fetus to recognized teratogens, such as thalidomide (Nowack, 1965) and rubella (Hill, Doll, Galloway, and Hughes, 1958). Applied to the present cases, however, the theory of maternal impression has obstacles as great as the normal explanation appears to have. First, in the 25 cases mentioned above, the subject's mother, although she may have heard of the death of the concerned deceased person, had no knowledge of that person's wounds. Second, this interpretation supposes that the mother not only modified the body of her unborn child with her thoughts, but after the child's birth influenced it to make statements and show behavior that it otherwise would not have done. No motive for such conduct can be discerned in most of the mothers (or fathers) of these subjects.

It is not my purpose to impose any interpretation of these cases on the readers of this article. Nor would I expect any reader to reach even a preliminary conclusion from the short summaries of cases that the brevity of this report entails. Instead, I hope that I have stimulated readers to examine the detailed reports of many cases that I am now in the process of publishing (Stevenson, forthcoming). "Originality and truth are found only in the details" (Stendhal, 1926).

Acknowledgements

I am grateful to Drs. Antonia Mills and Emily W. Cook for critical comments on drafts of this paper. Thanks are also due to the Bernstein Brothers Parapsychology and Health Foundation for the support of my research.

Correspondence and requests for reprints should be addressed to: Ian Stevenson, M.D., Division of Personality Studies, Box 152, Health Sciences Center, University of Virginia, Charlottesville, VA 22908

References

Cockayne, E. A. (1933). *Inherited abnormalities of the skin.* London: Oxford University Press.

Cook, E. W., Pasricha, S, Samararatne, G, Win Maung, & Stevenson, I. (1983). Review and analysis of "unsolved" cases of the reincarnation type: II. Comparison of features of solved and unsolved cases. *Journal of the American Society for Psychical Research,* 77, 115-135.

Dabney, W. C. (1890). Maternal impressions. In J. M. Keating (Ed.), *Cyclopaedia of the diseases of children,* Vol. 1, (pp. 191-216). Philadelphia: J. B. Lippincott.

Denaro, S. J. (1944). The inheritance of nevi. *Journal of Heredity,* 35, 215-18.

Fatteh, A. (1976). *Medicolegal investigation of gunshot wounds.* Philadelphia: J. B. Lippincott.

Frantz, C. H., & O'Rahilly, R.(1961). Congenital skeletal limb deficiencies. *Journal of Bone and Joint Surgery,* 43-A, 1202-24.

Gordon, I., & Shapiro, H. A. (1982). *Forensic medicine: A guide to principles.* (2nd ed.) London: Churchill Livingstone.

Hill, A. B., Doll, R., Galloway, T. M., & Hughes, J.P.W. (1958). Virus diseases in pregnancy and congenital defects. *British Journal of Preventive and Social Medicine,* 12, 1-7.

Maruri, C. A. (1961). *La herencia en dermatología.* (2nd ed.) Santander: Aldus, S.A. Artes Graficas.

Mills, A. (1989). A replication study: Three cases of children in northern India who are said to remember a previous life. *Journal of Scientific Exploration,* 3, 133-184.

Nelson, K., & Holmes, L. B. (1989). Malformations due to presumed spontaneous mutations in newborn infants. *New England Journal of Medicine,* 320, 19-23.

Nowack, E. (1965). Die sensible Phase bei der Thalidomid- Embryopathie. *Humangenetik,* 1, 516-36.

Pack, G. T., & Davis, J. (1956). Moles. *New York State Journal of Medicine,* 56, 3498-3506.

Pack, G. T., Lenson, N. & Gerber, D. M. (1952). Regional distribution of moles and melanomas. *AMA Archives of Surgery.* 65, 862-70.

Smith, D. W. (1982). *Recognizable patterns of human malformation.* (3rd ed.) Philadelphia: W. B. Saunders.

Spalteholz, W. (1943). *Hand atlas of human anatomy.* Translated by L. F. Barker. 7th English ed. Philadelphia: J.B. Lippincott.

Stendhal (1926). *Lucien Leuwen.* Paris: Librairie Ancienne Honoré Champion, 4, 169.

Stevenson, I. (1975). *Cases of the reincarnation type. I. Ten cases in India.* Charlottesville: University Press of Virginia.

Stevenson, I. (1983). American children who claim to remember previous lives. *Journal of Nervous and Mental Disease,* 171, 742-748.

Stevenson, I. (1987). *Children who remember previous lives.* Charlottesville: University Press of Virginia.

Stevenson, I. (1990). Phobias in children who claim to remember previous lives. *Journal of Scientific Exploration,* 4, 243-254.

Stevenson, I. (1992). A new look at maternal impressions: An analysis of 50 published cases and reports of two recent examples. *Journal of Scientific Exploration,* 6, 353-373.

Stevenson, I. (Forthcoming). *Birthmarks and birth defects: A contribution to their etiology.*

Williams, H. C., & Pembroke, A. C. (1988). Naevus of Jamaica. *Lancet,* ii, 915.

Wilson, J. G. (1973). *Environment and birth defects.* New York: Academic Press.

[5]

Near-death experience in survivors of cardiac arrest: a prospective study in the Netherlands

Pim van Lommel, Ruud van Wees, Vincent Meyers, Ingrid Elfferich

Summary

Background Some people report a near-death experience (NDE) after a life-threatening crisis. We aimed to establish the cause of this experience and assess factors that affected its frequency, depth, and content.

Methods In a prospective study, we included 344 consecutive cardiac patients who were successfully resuscitated after cardiac arrest in ten Dutch hospitals. We compared demographic, medical, pharmacological, and psychological data between patients who reported NDE and patients who did not (controls) after resuscitation. In a longitudinal study of life changes after NDE, we compared the groups 2 and 8 years later.

Findings 62 patients (18%) reported NDE, of whom 41 (12%) described a core experience. Occurrence of the experience was not associated with duration of cardiac arrest or unconsciousness, medication, or fear of death before cardiac arrest. Frequency of NDE was affected by how we defined NDE, the prospective nature of the research in older cardiac patients, age, surviving cardiac arrest in first myocardial infarction, more than one cardiopulmonary resuscitation (CPR) during stay in hospital, previous NDE, and memory problems after prolonged CPR. Depth of the experience was affected by sex, surviving CPR outside hospital, and fear before cardiac arrest. Significantly more patients who had an NDE, especially a deep experience, died within 30 days of CPR ($p<0.0001$). The process of transformation after NDE took several years, and differed from those of patients who survived cardiac arrest without NDE.

Interpretation We do not know why so few cardiac patients report NDE after CPR, although age plays a part. With a purely physiological explanation such as cerebral anoxia for the experience, most patients who have been clinically dead should report one.

Lancet 2001; **358:** 2039–45
See Commentary page 2010

Division of Cardiology, Hospital Rijnstate, Arnhem, Netherlands (P van Lommel MD); **Tilburg, Netherlands** (R van Wees PhD); **Nijmegen, Netherlands** (V Meyers PhD); **and Capelle a/d Ijssel, Netherlands** (I Elfferich PhD)

Correspondence to: Dr Pim van Lommel, Division of Cardiology, Hospital Rijnstate, PO Box 9555, 6800 TA Arnhem, Netherlands (e-mail: pimvanlommel@wanadoo.nl)

Introduction

Some people who have survived a life-threatening crisis report an extraordinary experience. Near-death experience (NDE) occurs with increasing frequency because of improved survival rates resulting from modern techniques of resuscitation. The content of NDE and the effects on patients seem similar worldwide, across all cultures and times. The subjective nature and absence of a frame of reference for this experience lead to individual, cultural, and religious factors determining the vocabulary used to describe and interpret the experience.[1]

NDE are reported in many circumstances: cardiac arrest in myocardial infarction (clinical death), shock in postpartum loss of blood or in perioperative complications, septic or anaphylactic shock, electrocution, coma resulting from traumatic brain damage, intracerebral haemorrhage or cerebral infarction, attempted suicide, near-drowning or asphyxia, and apnoea. Such experiences are also reported by patients with serious but not immediately life-threatening diseases, in those with serious depression, or without clear cause in fully conscious people. Similar experiences to near-death ones can occur during the terminal phase of illness, and are called deathbed visions. Identical experiences to NDE, so-called fear-death experiences, are mainly reported after situations in which death seemed unavoidable: serious traffic accidents, mountaineering accidents, or isolation such as with shipwreck.

Several theories on the origin of NDE have been proposed. Some think the experience is caused by physiological changes in the brain, such as brain cells dying as a result of cerebral anoxia.[2–4] Other theories encompass a psychological reaction to approaching death,[5] or a combination of such reaction and anoxia.[6] Such experiences could also be linked to a changing state of consciousness (transcendence), in which perception, cognitive functioning, emotion, and sense of identity function independently from normal body-linked waking consciousness.[1] People who have had an NDE are psychologically healthy, although some show non-pathological signs of dissociation.[7] Such people do not differ from controls with respect to age, sex, ethnic origin, religion, or degree of religious belief.[1]

Studies on NDE[1,3,8,9] have been retrospective and very selective with respect to patients. In retrospective studies, 5–10 years can elapse between occurrence of the experience and its investigation, which often prevents accurate assessment of physiological and pharmacological factors. In retrospective studies, between 43%[8] and 48%[1] of adults and up to 85% of children[10] who had a life-threatening illness were estimated to have had an NDE. A random investigation of more than 2000 Germans showed 4·3% to have had an NDE at a mean age of 22 years.[11] Differences in estimates of frequency and uncertainty as to causes of this experience result from varying definitions of the phenomenon, and from inadequate methods of

research.[12] Patients' transformational processes after an NDE are very similar[1,3,13–16] and encompass life-changing insight, heightened intuition, and disappearance of fear of death. Assimilation and acceptance of these changes is thought to take at least several years.[15]

We did a prospective study to calculate the frequency of NDE in patients after cardiac arrest (an objective critical medical situation), and establish factors that affected the frequency, content, and depth of the experience. We also did a longitudinal study to assess the effect of time, memory, and suppression mechanisms on the process of transformation after NDE, and to reaffirm the content and allow further study of the experience. We also proposed to reassess theories on the cause and content of NDE.

Methods

Patients

We included consecutive patients who were successfully resuscitated in coronary care units in ten Dutch hospitals during a research period varying between hospitals from 4 months to nearly 4 years (1988–92). The research period varied because of the requirement that all consecutive patients who had undergone successful cardiopulmonary resuscitation (CPR) were included. If this standard was not met we ended research in that hospital. All patients had been clinically dead, which we established mainly by electrocardiogram records. All patients gave written informed consent. We obtained ethics committee approval.

Procedures

We defined NDE as the reported memory of all impressions during a special state of consciousness, including specific elements such as out-of-body experience, pleasant feelings, and seeing a tunnel, a light, deceased relatives, or a life review. We defined clinical death as a period of unconsciousness caused by insufficient blood supply to the brain because of inadequate blood circulation, breathing, or both. If, in this situation, CPR is not started within 5–10 min, irreparable damage is done to the brain and the patient will die.

We did a short standardised interview with sufficiently well patients within a few days of resuscitation. We asked whether patients recollected the period of unconsciousness, and what they recalled. Three researchers coded the experiences according to the weighted core experience index.[1] In this scoring system, depth of NDE is measured with weighted scores assigned to elements of the content of the experience. Scores between 1 and 5 denote superficial NDE, but we included these events because all patients underwent transformational changes as well. Scores of 6 or more denote core experiences, and scores of 10 or greater are deep experiences. We also recorded date of cardiac arrest, date of interview, sex, age, religion, standard of education reached, whether the patient had previously experienced NDE, previously heard of NDE, whether CPR took place inside or outside hospital, previous myocardial infarction, and how many times the patient had been resuscitated during their stay in hospital. We estimated duration of circulatory arrest and unconsciousness, and noted whether artificial respiration by intubation took place. We also recorded type and dose of drugs before, during, and after the crisis, and assessed possible memory problems at interview after lengthy or difficult resuscitation. We classed patients resuscitated during electrophysiological stimulation separately.

We did standardised and taped interviews with participants a mean of 2 years after CPR. Patients also completed a life-change inventory.[16] The questionnaire addressed self-image, concern with others, materialism and social issues, religious beliefs and spirituality, and attitude towards death. Participants answered 34 questions with a five-point scale indicating whether and to what degree they had changed. After 8 years, surviving patients and their partners were interviewed again with the life-change inventory, and also completed a medical and psychological questionnaire for cardiac patients (from the Dutch Heart Foundation), the Utrecht coping list, the sense of coherence inquiry, and a scale for depression. These extra questionnaires were deemed necessary for qualitative analysis because of the reduced number of respondents who survived to 8 years follow-up. Our control group consisted of resuscitated patients who had not reported an NDE. We matched controls with patients who had had an NDE by age, sex, and time interval between CPR and the second and third interviews.

Statistical analysis

We assessed causal factors for NDE with the Pearson χ^2 test for categorical and t test for ratio-scaled factors. Factors affecting depth of NDE were analysed with the Mann-Whitney test for categorical factors, and with Spearman's coefficient of rank correlation for ratio-scaled factors. Links between NDE and altered scores for questions from the life-change inventory were assessed with the Mann-Whitney test. The sums of the individual scores were used to compare the responses to the life-change inventory in the second and third interview. Because few causes or relations exist for NDE, the null hypotheses are the absence of factors. Hence, all tests were two-tailed with significance shown by p values less than 0·05.

Results

Patients

We included 344 patients who had undergone 509 successful resuscitations. Mean age at resuscitation was 62·2 years (SD 12·2), and ranged from 26 to 92 years. 251 patients were men (73%) and 93 were women (27%). Women were significantly older than men (66 *vs* 61 years, p=0·005).The ratio of men to women was 57/43 for those older than 70 years, whereas at younger ages it was 80/20. 14 (4%) patients had had a previous NDE. We interviewed 248 (74%) patients within 5 days after CPR. Some demographic questions from the first interview had too many values missing for reliable statistical analysis, so data from the second interview were used. Of the 74 patients whom we interviewed at 2-year follow-up, 42 (57%) had previously heard of NDE, 53 (72%) were religious, 25 (34%) had left education aged 12 years, and 49 (66%) had been educated until aged at least 16 years.

296 (86%) of all 344 patients had had a first myocardial infarction and 48 (14%) had undergone more than one infarction. Nearly all patients with acute myocardial infarction were treated with fentanyl, a synthetic opiod antagonist; thalamonal, a combined preparation of fentanyl with dehydrobenzperidol that has an antipsychotic and sedative effect; or both. 45 (13%) patients also received sedative drugs such as diazepam or oxazepam, and 38 (11%) were given strong sedatives such as midazolam (for intubation), or haloperidol for cerebral unrest during or after long-lasting unconsciousness.

	WCEI score*	n
A No memory	0	282 (82%)
B Some recollection	1–5	21 (6%)
C Moderately deep NDE	6–9	18 (5%)
D Deep NDE	10–14	17 (5%)
E Very deep NDE	15–19	6 (2%)

WCEI=weighted core experience index. NDE=near-death experience. *A=no NDE, B=superficial NDE, C/D/E=core NDE

Table 1: **Distribution of the 344 patients in five WCEI classes***

234 (68%) patients were successfully resuscitated within hospital. 190 (81%) of these patients were resuscitated within 2 min of circulatory arrest, and unconsciousness lasted less than 5 min in 187 (80%). 30 patients were resuscitated during electrophysiological stimulation; these patients all underwent less than 1 min of circulatory arrest and less than 2 min of unconsciousness. This group were only given 5 mg of diazepam about 1 h before electrophysiological stimulation.

101 (29%) patients survived CPR outside hospital, and nine (3%) were resuscitated both within and outside hospital. Of these 110 patients, 88 (80%) had more than 2 min of circulatory arrest, and 62 (56%) were unconscious for more than 10 min. All people with brief cardiac arrest and who were resuscitated outside hospital were resuscitated in an ambulance. Only 12 (9%) patients survived a circulatory arrest that lasted longer than 10 min. 36% (123) of all patients were unconsciousness for longer than 60 min, 37 of these patients needed artificial respiration through intubation. Intubated patients received high doses of strong sedatives and were interviewed later than other patients; most were still in a weakened physical condition at the time of first interview and 24 showed memory defects. Significantly more younger than older patients survived long-lasting unconsciousness following difficult CPR (p=0·005).

Prospective findings

62 (18%) patients reported some recollection of the time of clinical death (table 1). Of these patients, 21 (6% of total) had a superficial NDE and 41 (12%) had a core experience. 23 of the core group (7% of total) reported a deep or very deep NDE. Therefore, of 509 resuscitations, 12% resulted in NDE and 8% in core experiences. Table 2 shows the frequencies of ten elements of NDE.[1] No patients reported distressing or frightening NDE.

During the pilot phase in one of the hospitals, a coronary-care-unit nurse reported a veridical out-of-body experience of a resuscitated patient:

"During a night shift an ambulance brings in a 44-year-old cyanotic, comatose man into the coronary care unit. He had been found about an hour before in a meadow by passers-by. After admission, he receives artificial respiration without intubation, while heart massage and defibrillation are also applied. When we want to intubate the patient, he turns out to have dentures in his mouth. I remove these upper dentures and put them onto the 'crash car'. Meanwhile, we continue extensive CPR. After about an hour and a half the patient has sufficient heart rhythm and blood pressure, but he is still ventilated and intubated, and he is still comatose. He is transferred to the intensive care unit to continue the necessary artificial respiration. Only after more than a week do I meet again with the patient, who is by now back on the cardiac ward. I distribute his medication. The moment he sees me he says: 'Oh, that nurse knows where my dentures are'. I am very surprised. Then he elucidates: 'Yes, you were there when I was brought into hospital and you took my dentures out of my mouth and put them onto that car, it had all these bottles on it and there was this sliding drawer underneath and there you put my teeth.' I was especially amazed because I remembered this happening while the man was in deep coma and in the process of CPR. When I asked further, it appeared the man had seen himself lying in bed, that he had perceived from above how nurses and doctors had been busy with CPR. He was also able to describe correctly and in detail the small room in which he had been resuscitated as well as the appearance of those present like myself. At the time that he observed the situation he had been very much afraid that we would stop CPR and that he would die. And it is true that we had been very negative about the patient's prognosis due to his very poor medical condition when admitted. The patient tells me that he desperately and unsuccessfully tried to make it clear to us that he was still alive and that we should continue CPR. He is deeply impressed by his experience and says he is no longer afraid of death. 4 weeks later he left hospital as a healthy man."

Elements of NDE[1]	Frequency (n=62)
1 Awareness of being dead	31 (50%)
2 Positive emotions	35 (56%)
3 Out of body experience	15 (24%)
4 Moving through a tunnel	19 (31%)
5 Communication with light	14 (23%)
6 Observation of colours	14 (23%)
7 Observation of a celestial landscape	18 (29%)
8 Meeting with deceased persons	20 (32%)
9 Life review	8 (13%)
10 Presence of border	5 (8%)

NDE=near-death experience

Table 2: **Frequency of ten elements of NDE**

Table 3 shows relations between demographic, medical, pharmacological, and psychological factors and the frequency and depth of NDE. No medical, pharmacological, or psychological factor affected the frequency of the experience. People younger than 60 years had NDE more often than older people (p=0·012), and women, who were significantly older than men, had more frequent deep experiences than men (p=0·011) (table 3). Increased frequency of experiences in patients who survived cardiac arrest in first myocardial infarction, and deeper experiences in patients who survived CPR outside hospital could have resulted from differences in age. Both these groups of patients were younger than other patients, though the age differences were not significant (p=0·05 and 0·07, respectively).

Lengthy CPR can sometimes induce loss of memory and patients thus affected reported significantly fewer NDEs than others (table 3). No relation was found between frequency of NDE and the time between CPR and the first interview (range 1–70 days). Mortality during or shortly after stay in hospital in patients who had an NDE was significantly higher than in patients who did not report an NDE (13/62 patients [21%] *vs* 24/282 [9%], p=0·008), and this difference was even more marked in patients who reported a deep experience (10/23 [43%] *vs* 24/282 [9%], p<0·0001).

Longitudinal findings

At 2-year follow-up, 19 of the 62 patients with NDE had died and six refused to be interviewed. Thus, we were able to interview 37 patients for the second time. All

	Frequency of NDE			Depth of NDE (n=62)
	NDE (n=62)	No NDE (n=282)	p	
Categorical factors				
Demographic				
Women	13 (21%)	80 (28%)	NS	0·011
Age* <60 years	32 (52%)	96 (34%)	0·012	NS
Religion† (yes)	26 (70%)	27 (73%)	NS	NS
Education†‡ Elementary	10 (27%)	15 (43%)	NS	NS
Medical				
Intubation	6 (10%)	31 (11%)	NS	NS
Electrophysiological stimulation	8 (13%)	22 (8%)	NS	NS
First myocardial infarction	60 (97%)	236 (84%)	0·013	NS
CPR outside hospital§	13 (21%)	88 (32%)	NS	0 027
Memory defect after lengthy CPR	1 (2%)	40 (14%)	0·011	NS
Death within 30 days	13 (21%)	24 (9%)	0·008	0·017
Pharmacological				
Extra medication	17 (27%)	70 (25%)	NS	NS
Psychological				
Fear before CPR†§	4 (13%)	2 (6%)	NS	0·045
Previous NDE	6 (10%)	8 (3%)	0·035	NS
Foreknowledge of NDE†	22 (60%)	20 (54%)	NS	NS
Ratio-scaled factors				
Demographic				
Age (mean [SD], years)*	58·8 (13 4)	63·5 (11·8)	0·006	NS
Medical				
Duration of cardiac arrest (mean [SD], min)	4·0 (5·2)	3·7 (3·9)	NS	NS
Duration of unconsciousness (mean [SD], min)	66·1 (269·5)	118 3 (355·5)	NS	NS
Number of CPRs (SD)	2·1 (2 5)	1 4 (1·2)	0·029	NS

Data are number (%) unless otherwise indicated CPR=cardiopulmonary resuscitation. NS=not significant (p>0 05) *3 missing values †n=74 (data from 2nd interview, 35 NDE, 39 no NDE) ‡2 missing values §10 missing values

Table 3: **Factors affecting frequency and depth of near-death experience (NDE)**

patients were able to retell their experience almost exactly. Of the 17 patients who had low scores in the first interview (superficial NDE), seven had unchanged low scores, and four probably had, in retrospect, an NDE that consisted only of positive emotions (score 1). Six patients had not in fact had an NDE after all, which was probably because of our wide definition of NDE at the first interview.

We selected a control group, matched for age, sex, and time since cardiac arrest, from the 282 patients who had not had NDE. We contacted 75 of these patients to obtain 37 survivors who agreed to be interviewed. Two controls reported an NDE consisting only of positive emotions, and two a core experience. The first interview after CPR might have been too soon for these four patients (1% of total) to remember their NDE, or to be willing or able to describe the experience. We were therefore able to interview 35 patients who had had an affirmed NDE, and 39 patients who had not.

Only six of the 74 patients that we interviewed at 2 years said they were afraid before CPR (table 3). Four of these six had deep NDE (p=0·045, table 3). Most patients were not afraid before CPR, as the arrest happened too suddenly and unexpectedly to allow time for fear.

Significant differences in answers to 13 of the 34 items in the life-change inventory between people with and without an NDE are shown in table 4. For instance, people who had NDE had a significant increase in belief in an afterlife and decrease in fear of death compared with people who had not had this experience. Depth of

Life-change inventory questionnaire	p
Social attitude	
Showing own feelings	0 034
Acceptance of others*	0 012
More loving, empathic*	0 002
Understanding others*	0 003
Involvement in family*	0·008
Religious attitude	
Understand purpose of life*	0·020
Sense inner meaning of life*	0·028
Interest in spirituality*	0 035
Attitude to death	
Fear of death*	0·009
Belief in life after death*	0·007
Others	
Interest in meaning of life	0 020
Understanding oneself	0 019
Appreciation of ordinary things	0·0001

NDE=near-death experience 35 patients had NDE, 39 had not had NDE 1 value missing for patients with NDE in all categories, *2 values missing for patients with NDE (ie, n=33)

Table 4: **Significant differences in life-change inventory-scores[16] of patients with and without NDE at 2-year follow-up**

NDE was linked to high scores in spiritual items such as interest in the meaning of one's own life, and social items such as showing love and accepting others. The 13 patients who had superficial NDE underwent the same specific transformational changes as those who had a core experience.

8-year follow-up included 23 patients with an NDE that had been affirmed at 2-year follow-up. 11 patients had died and one could not be interviewed. Patients could still recall their NDE almost exactly. Of the patients without an NDE at 2-year follow-up, 20 had died and four patients could not be interviewed (for reasons such as dementia and long stay in hospital), which left 15 patients without an NDE to take part in the third interview.

All patients, including those who did not have NDE, had gone through a positive change and were more self-assured, socially aware, and religious than before. Also,

Life-change inventory questionnaire	2-year follow-up		8-year follow-up	
	NDE (n=23)	no NDE (n=15)	NDE (n=23)	no NDE (n=15)
Social attitude				
Showing own feelings	42	16	78	58
Acceptance of others	42	16	78	41
More loving, empathic	52	25	68	50
Understanding others	36	8	73	75
Involvement in family	47	33	78	58
Religious attitude				
Understand purpose of life	52	33	57	66
Sense inner meaning of life	52	25	57	25
Interest in spirituality	15	-8	42	-41
Attitude to death				
Fear of death	-47	-16	-63	-41
Belief in life after death	36	16	42	16
Others				
Interest in meaning of life	52	33	89	66
Understanding oneself	58	8	63	58
Appreciation of ordinary things	78	41	84	50

NDE=near-death experience The sums of all individual scores per item are reported in the same 38 patients who had both follow-up interviews Participants responded in a five-point scale indicating whether and to what degree they had changed strongly increased (+2), somewhat increased (+1), no change (0), somewhat decreased (-1), and strongly decreased (-2) Only in the reported 13 (of 34) items in this table were significant differences found in life-change scores in the interview after 2 years (table 4)

Table 5: **Total sum of individual life-change inventory scores[16] of patients at 2-year and 8-year follow-up**

people who did not have NDE had become more emotionally affected, and in some, fear of death had decreased more than at 2-year follow-up. Their interest in spirituality had strongly decreased. Most patients who did not have NDE did not believe in a life after death at 2-year or 8-year follow-up (table 5). People with NDE had a much more complex coping process: they had become more emotionally vulnerable and empathic, and often there was evidence of increased intuitive feelings. Most of this group did not show any fear of death and strongly believed in an afterlife. Positive changes were more apparent at 8 years than at 2 years of follow-up.

Discussion

Our results show that medical factors cannot account for occurrence of NDE; although all patients had been clinically dead, most did not have NDE. Furthermore, seriousness of the crisis was not related to occurrence or depth of the experience. If purely physiological factors resulting from cerebral anoxia caused NDE, most of our patients should have had this experience. Patients' medication was also unrelated to frequency of NDE. Psychological factors are unlikely to be important as fear was not associated with NDE.

The 18% frequency of NDE that we noted is lower than reported in retrospective studies,[1,8] which could be because our prospective study design prevented self-selection of patients. Our frequency of NDE is low despite our wide definition of the experience. Only 12% of patients had a core NDE, and this figure might be an overestimate. When we analysed our results, we noted that one hospital that participated in the study for nearly 4 years, and from which 137 patients were included, reported a significantly ($p=0\cdot01$) lower percentage of NDE (8%), and significantly ($p=0\cdot05$) fewer deep experiences. Therefore, possibly some selection of patients occurred in the other hospitals, which sometimes only took part for a few months. In a prospective study[17] with the same design as ours, 6% of 63 survivors of cardiac arrest reported a core experience, and another 5% had memories with features of an NDE (low score in our study); thus, with our wide definition of the experience, 11% of these patients reported an NDE. Therefore, true frequency of the experience is likely to be about 10%, or 5% if based on number of resuscitations rather than number of resuscitated patients. Patients who survive several CPRs in hospital have a significantly higher chance of NDE (table 3).

We noted that the frequency of NDE was higher in people younger than 60 years than in older people. In other studies, mean age at NDE is lower than our estimate (62·2 years) and the frequency of the experience is higher. Morse[10] saw 85% NDE in children, Ring[1] noted 48% NDE in people with a mean age of 37 years, and Sabom[8] saw 43% NDE in people with a mean age of 49 years; thus, age and the frequency of the experience seem to be associated. Other retrospective studies have noted a younger mean age for NDE: 32 years,[9] 29 years,[6] and 22 years.[11] Cardiac arrest was the cause of the experience in most patients in Sabom's[8] study, whereas this was the case in only a low percentage of patients in other work. We saw that people surviving CPR outside hospital (who underwent deeper NDE than other patients) tended to be younger, as were those who survived cardiac arrest in a first myocardial infarction (more frequent NDE), which indicates that age was probably decisive in the significant relation noted with those factors.

In a study of mortality in patients after resuscitation outside hospital,[18] chances of survival increased in people younger than 60 years and in those undergoing first myocardial infarction, which corresponds with our findings. Older people have a smaller chance of cerebral recovery after difficult and complicated resuscitation after cardiac arrest. Younger patients have a better chance of surviving a cardiac arrest, and thus, to describe their experience. In a study of 11 patients after CPR, the person that had an NDE was significantly younger than other patients who did not have such an experience.[19] Greyson[7] also noted a higher frequency of NDE and significantly deeper experiences at younger ages, as did Ring.[1]

Good short-term memory seems to be essential for remembering NDE. Patients with memory defects after prolonged resuscitation reported fewer experiences than other patients in our study. Forgetting or repressing such experiences in the first days after CPR was unlikely to have occurred in the remaining patients, because no relation was found between frequency of NDE and date of first interview. However, at 2-year follow-up, two patients remembered a core NDE and two an NDE that consisted of only positive emotions that they had not reported shortly after CPR, presumably because of memory defects at that time. It is remarkable that people could recall their NDE almost exactly after 2 and 8 years.

Unlike our results, an inverse correlation between foreknowledge and frequency of NDE has been shown.[1,8] Our finding that women have deeper experiences than men has been confirmed in two other studies,[1,7] although in one,[7] only in those cases in which women had an NDE resulting from disease.

The elements of NDE that we noted (table 2) correspond with those in other studies based on Ring's[1] classification. Greyson[20] constructed the NDE scale differently to Ring,[1] but both scoring systems are strongly correlated ($r=0\cdot90$). Yet, reliable comparisons are nearly impossible between retrospective studies that included selection of patients, unreliable medical records, and used different criteria for NDE,[12] and our prospective study.

Our longitudinal follow-up research into transformational processes after NDE confirms the transformation described by many others.[1-3,8,10,13-16,21] Several of these investigations included a control group to enable study of differences in transformation,[14] but in our research, patients were interviewed three times during 8 years, with a matched control group. Our findings show that this process of change after NDE tends to take several years to consolidate. Presumably, besides possible internal psychological processes, one reason for this has to do with society's negative response to NDE, which leads individuals to deny or suppress their experience for fear of rejection or ridicule. Thus, social conditioning causes NDE to be traumatic, although in itself it is not a psychotraumatic experience. As a result, the effects of the experience can be delayed for years, and only gradually and with difficulty is an NDE accepted and integrated. Furthermore, the longlasting transformational effects of an experience that lasts for only a few minutes of cardiac arrest is a surprising and unexpected finding.

One limitation of our study is that our study group were all Dutch cardiac patients, who were generally older than groups in other studies. Therefore, our frequency of NDE might not be representative of all cases—eg, a higher frequency could be expected with

younger samples, or rates might vary in other populations. Also, the rates for NDE could differ in people who survive near-death episodes that come about by different causes, such as near drowning, near fatal car crashes with cerebral trauma, and electrocution. However, rigorous prospective studies would be almost impossible in many such cases.

Several theories have been proposed to explain NDE. We did not show that psychological, neurophysiological, or physiological factors caused these experiences after cardiac arrest. Sabom[22] mentions a young American woman who had complications during brain surgery for a cerebral aneurysm. The EEG of her cortex and brainstem had become totally flat. After the operation, which was eventually successful, this patient proved to have had a very deep NDE, including an out-of-body experience, with subsequently verified observations during the period of the flat EEG.

And yet, neurophysiological processes must play some part in NDE. Similar experiences can be induced through electrical stimulation of the temporal lobe (and hence of the hippocampus) during neurosurgery for epilepsy,[23] with high carbon dioxide levels (hypercarbia),[24] and in decreased cerebral perfusion resulting in local cerebral hypoxia as in rapid acceleration during training of fighter pilots,[25] or as in hyperventilation followed by valsalva manoeuvre.[4] Ketamine-induced experiences resulting from blockage of the NMDA receptor,[26] and the role of endorphin, serotonin, and enkephalin have also been mentioned,[27] as have near-death-like experiences after the use of LSD,[28] psilocarpine, and mescaline.[21] These induced experiences can consist of unconsciousness, out-of-body experiences, and perception of light or flashes of recollection from the past. These recollections, however, consist of fragmented and random memories unlike the panoramic life-review that can occur in NDE. Further, transformational processes with changing life-insight and disappearance of fear of death are rarely reported after induced experiences.

Thus, induced experiences are not identical to NDE, and so, besides age, an unknown mechanism causes NDE by stimulation of neurophysiological and neurohumoral processes at a subcellular level in the brain in only a few cases during a critical situation such as clinical death. These processes might also determine whether the experience reaches consciousness and can be recollected.

With lack of evidence for any other theories for NDE, the thus far assumed, but never proven, concept that consciousness and memories are localised in the brain should be discussed. How could a clear consciousness outside one's body be experienced at the moment that the brain no longer functions during a period of clinical death with flat EEG?[22] Also, in cardiac arrest the EEG usually becomes flat in most cases within about 10 s from onset of syncope.[29,30] Furthermore, blind people have described veridical perception during out-of-body experiences at the time of this experience.[31] NDE pushes at the limits of medical ideas about the range of human consciousness and the mind-brain relation.

Another theory holds that NDE might be a changing state of consciousness (transcendence), in which identity, cognition, and emotion function independently from the unconscious body, but retain the possibility of non-sensory perception.[7,8,22,28,31]

Research should be concentrated on the effort to explain scientifically the occurrence and content of NDE. Research should be focused on certain specific elements of NDE, such as out-of-body experiences and other verifiable aspects. Finally, the theory and background of transcendence should be included as a part of an explanatory framework for these experiences.

Contributors

Pim van Lommel coordinated the first interviews and was responsible for collecting all demographic, medical, and pharmacological data. Pim van Lommel, Ruud van Wees, and Vincent Meyers rated the first interview. Ruud van Wees and Vincent Meyers coordinated the second interviews. Ruud van Wees did statistical analysis of the first and second interviews. Ingrid Elfferich did the third interviews and analysed these results.

Acknowledgments

We thank nursing and medical staff of the hospitals involved in the research; volunteers of the International Association of Near Death Studies; IANDS-Netherlands; Merkawah Foundation for arranging interviews, and typing the second and third interviews; Martin Meyers for help with translation; and Kenneth Ring and Bruce Greyson for review of the article.

References

1 Ring K. Life at death. A scientific investigation of the near-death experience. New York: Coward McCann and Geoghenan, 1980.
2 Blackmore S. Dying to live: science and the near-death experience. London: Grafton—an imprint of Harper Collins Publishers, 1993.
3 Morse M. Transformed by the light. New York: Villard Books, 1990.
4 Lempert T, Bauer M, Schmidt D. Syncope and near-death experience. *Lancet* 1994; **344:** 829–30.
5 Appelby L. Near-death experience: analogous to other stress induced physiological phenomena. *BMJ* 1989; **298:** 976–77.
6 Owens JE, Cook EW, Stevenson I. Features of "near-death experience" in relation to whether or not patients were near death. *Lancet* 1990; **336:** 1175–77.
7 Greyson B. Dissociation in people who have near-death experiences. out of their bodies or out of their minds? *Lancet* 2000; **355:** 460–63.
8 Sabom MB. Recollections of death: a medical investigation. New York: Harper and Row, 1982.
9 Greyson B. Varieties of near-death experience. *Psychiatry* 1993; **56:** 390–99.
10 Morse M. Parting visions: a new scientific paradigm. In: Bailey LW, Yates J, eds. The near-death experience: a reader. New York and London: Routledge, 1996: 299–318.
11 Schmied I, Knoblaub H, Schnettler B. Todesnaheerfahrungen in Ost- und Westdeutschland—eine empirische Untersuchung. In: Knoblaub H, Soeffner HG, eds. Todesnahe: interdisziplinäre Zugänge zu einem außergewohnlichen Phänomen. Konstanz: Universitatsverlag Konstanz, 1999: 217–50.
12 Greyson B. The incidence of near-death experiences. *Med Psychiatry* 1998; **1:** 92–99.
13 Roberts G, Owen J. The near-death experience. *Br J Psychiatry* 1988; **153:** 607–17.
14 Groth-Marnat G, Summers R. Altered beliefs, attitudes and behaviors following near-death experiences. *J Hum Psychol* 1998; **38:** 110–25.
15 Atwater PMH. Coming back to life: the after-effects of the near-death experience. New York: Dodd, Mead and Company, 1988.
16 Ring K. Heading towards omega: in search of the meaning of the near-death experience. New York: Quill William Morrow, 1984.
17 Parnia S, Waller DG, Yeates R, Fenwick P. A qualitative and quantitative study of the incidence, features and aetiology of near death experiences in cardiac arrest survivors. *Resuscitation* 2001; **48:** 149–56.
18 Dickey W, Adgey AAJ. Mortality within hospital after resuscitation from ventricular fibrillation outside hospital. *Br Heart J* 1992; **67:** 334–38.
19 Schoenbeck SB, Hocutt GD. Near-death experiences in patients undergoing cardio-pulmonary resuscitation *J Near-Death Studies* 1991; **9:** 211–18.
20 Greyson B. The near-death experience scale: construction, reliability and validity. *J Nervous Mental Dis* 1982; **171:** 369–75.

21 Schroter-Kunhardt M. Nah—Todeserfahrungen aus psychiatrisch-neurologischer Sicht. In: Knoblaub H, Soeffner HG, eds Todesnähe: interdisziplinare Zugänge zu einem außergewohnlichen Phanomen. Konstanz. Universitätsverlag Konstanz, 1999: 65–99

22 Sabom MB. Light and death: one doctors fascinating account of near-death experiences. Michigan: Zondervan Publishing House, 1998: 37–52.

23 Penfield W. The excitable cortex in conscious man. Liverpool: Liverpool University Press, 1958.

24 Meduna LT. Carbon dioxide therapy: a neuropsychological treatment of nervous disorders. Springfield. Charles C Thomas, 1950.

25 Whinnery JE, Whinnery AM Acceleration-induced loss of consciousness. *Arch Neurol* 1990; **47:** 764–76.

26 Jansen K. Neuroscience, ketamine and the near-death experience: the role of glutamate and the NMDA-receptor In: Bailey LW, Yates J, eds. The near-death experience: a reader. New York and London: Routledge, 1996: 265–82.

27 Greyson B. Biological aspects of near-death experiences. *Perspect Biol Med* 1998; **42:** 14–32.

28 Grof S, Halifax J. The human encounter with death. New York: Dutton, 1977.

29 Clute HL, Levy WJ. Electroencephalographic changes during brief cardiac arrest in humans. *Anesthesiology* 1990; **73:** 821–25.

30 Aminoff MJ, Scheinman MM, Griffing JC, Herre JM. Electrocerebral accompaniments of syncope associated with malignant ventricular arrhythmias. *Ann Intern Med* 1988, **108:** 791–96.

31 Ring K, Cooper S Mindsight. near-death and out-of-body experiences in the blind. Palo Alto: William James Center for Consciousness Studies, 1999.

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[6]

An investigation into alleged 'hauntings'

Richard Wiseman[1]*, Caroline Watt[2], Paul Stevens[2], Emma Greening[1] and Ciarán O'Keeffe[1]
[1]University of Hertfordshire, UK
[2]University of Edinburgh, UK

In cases of alleged hauntings, a large number of seemingly trustworthy witnesses consistently report experiencing unusual phenomena (e.g. apparitions, sudden changes in temperature, a strong sense of presence) in certain locations. The two studies reported here explored the psychological mechanisms that underlie this apparent evidence of 'ghostly' activity. The experiments took place at two locations that have a considerable reputation for being haunted—Hampton Court Palace (Surrey, England) and the South Bridge Vaults (Edinburgh, Scotland). Both studies involved participants walking around these locations and reporting where they experienced unusual phenomena. Results revealed significantly more reports of unusual experiences in areas that had a reputation for being haunted. This effect was not related to participants' prior knowledge about the reputation of these areas. However, the location of participants' experiences correlated significantly with various environmental factors, including, for example, the variance of local magnetic fields and lighting levels. These findings strongly suggest that alleged hauntings may not necessarily represent evidence for 'ghostly' activity, but could be, at least in part, the result of people responding to 'normal' factors in their surroundings.

Recent polls reveal that approximately 38% of Americans believe that ghosts exist (Gallup, 2001), and 13% report having experienced one (MORI, 1998). Such experiences involve a diverse range of phenomena, including apparitions, unusual odours, sudden changes in temperature and a strong sense of presence (Lange, Houran, Harte, & Havens, 1996). In a relatively small number of cases, witnesses consistently report these experiences in certain locations, often giving rise to the belief that these places are 'haunted'. The best of these cases appear evidentially impressive, sometimes lasting several years and involving a large number of seemingly trustworthy witnesses reporting unusual phenomena in the same 'haunted' areas (for further information see Gauld & Cornell, 1979; Houran & Lange, 2001; Irwin, 1999; McCue, 2002). Many of these alleged

*Requests for reprints should be addressed to Dr Richard Wiseman, University of Hertfordshire, Hatfield Campus, College Lane, Hatfield, Herts AL10 9AB, UK (e-mail: r.wiseman@herts.ac.uk).

hauntings have been described in several best-selling books on the paranormal, and reported on both television and radio (see e.g. Auerbach, 1986).

These high-profile claims have been the subject of very little well-controlled, systematic, research. This is unfortunate, in part, because media reportage of many of these cases exerts a major influence over the public's belief in the paranormal (National Science Board, 2000). In addition, such work clearly has the potential to contribute to our theoretical understanding of how certain psychological and psychophysiological phenomena (including e.g. hallucination, suggestion and response to subtle environmental stimuli) operate in unusual, but naturalistic, settings (see e.g. Houran & Lange, 1996: Houran & Williams, 1998; Lange & Houran, 1997). The work also could contribute to applied research into several important, and often controversial, areas, including e.g. contagious psychogenic illness, sick building syndrome and other forms of alleged 'environmental illness' (Lundberg, 1998). The present article addresses these issues by outlining the first investigations into two internationally known cases of alleged hauntings.

Experiment 1 took place at Hampton Court Palace. This royal palace was home to many British monarchs for over 500 years, and it is now a popular historical attraction. The palace is also frequently referred to as 'one of the most haunted places in England' (see e.g. Guiley, 1994; Law, 1918: Underwood, 1971), and allegedly contains the ghost of Catherine Howard, the fifth wife of Henry VIII. Fifteen months after her marriage to the King in 1540, Catherine Howard was found guilty of adultery and sentenced to death (Thurley, 1996). Legend suggests that upon hearing the news, Catherine Howard ran to the King to plead for her life, but was dragged back along a section of the Palace now known as 'the Haunted Gallery' (Guiley, 1994; Underwood, 1971). By the turn of the century, the Gallery had become associated with various unusual experiences, including sightings of a 'woman in white' and reports of inexplicable screams (Law, 1918). Since then, visitors to the Gallery have reported other 'ghostly' phenomena, including a strong sense of presence, a feeling of dizziness and sudden changes in temperature (Franklin, 1998). The Haunted Gallery is not the only part of Hampton Court Palace associated with such phenomena, with visitors and staff reporting similar experiences in other areas of the building, including an area known as the Georgian Rooms (Franklin, 1998). Information about the reputation of the Haunted Gallery is widely available to the public, but specific information about the location of experiences in areas such as the Georgian Rooms is not widely available.

Experiment 2 was carried out in part of the South Bridge Vaults in Edinburgh, Scotland. Edinburgh's South Bridge was constructed in the late eighteenth century to ease transportation problems in the city. The Bridge consisted of 19 huge stone arches supporting a wide road lined with several three storey buildings. A series of 'vaults' (i.e. small chambers, rooms and corridors) were built into the Bridge's arches to house workshops, storage areas and accommodation for the poor (Henderson, 1999). However, ineffective waterproofing and overcrowding meant that by the mid-nineteenth century the vaults had degenerated into a disease-ridden slum. The area was abandoned during the late nineteenth century, but rediscovered and opened for public tours in 1996. During some of these tours, both members of the public and guides have experienced many unusual phenomena, including, for example, a strong sense of presence, several apparitions and 'ghostly' footsteps (Wilson, Brogan, & Hollinrake, 1999). As a result, the vaults have acquired an international reputation for being one of the most haunted parts of Scotland's capital city. The public has relatively easy access to

general information about haunt experiences in the vaults, but specific information about the location of particular experiences is not widely available.

EXPERIMENT 1

The first part of Expt 1 examined whether participants would report a disproportionately large number of unusual experiences in apparently 'haunted' areas of the Haunted Gallery and the Georgian Rooms at Hampton Court Palace. Prior to the study, Ian Franklin (IF), a warder at the palace, catalogued many of the reports of unusual phenomena associated with the building. IF reviewed this material and identified areas where people had consistently reported unusual phenomena in both the Haunted Gallery and the Georgian Rooms. The areas identified were classified as 'haunted' whilst the remaining areas were classified as 'controls'. The investigators were blind to these classifications until all data collection had been completed.

Groups of participants walked around either the Haunted Gallery or the Georgian Rooms, and reported if they experienced any unusual phenomena. Participants reporting such phenomena marked the locations of their experiences on a floorplan. It was predicted that the percentage of experiences reported in the 'haunted' areas in both locations would be significantly above chance.

Some researchers have argued that the witnesses involved in alleged hauntings may have had prior knowledge about which parts of a building were 'haunted', and that this may be responsible for them reporting a disproportionately large number of unusual experiences in these areas. There are several ways in which this may happen. For example, witnesses' prior knowledge about a 'haunted' area may cause them to assign special significance to any unusual phenomenon experienced in that area, therefore increasing the likelihood of them telling others about their experience. Alternatively, such information may have increased witnesses' anxiety levels when entering these areas, and this, in turn, may have resulted in the witnesses experiencing mild psychosomatic and hallucinatory phenomena. A second part of Expt 1 evaluated whether any disproportionate reporting of unusual experiences in 'haunted' areas would be due to participants' prior knowledge about previous reports of 'ghostly' activity. Prior to visiting either the Haunted Gallery or the Georgian Rooms, participants rated the degree to which they knew where in these locations people had experienced 'ghostly' phenomena in the past. The 'prior knowledge' hypothesis predicted that participants indicating a high level of prior knowledge would report a greater percentage of experiences in the 'haunted' areas than those indicating a low level of prior knowledge. Of course, participants can only report on their prior *conscious* knowledge. It is theoretically possible that participants may be influenced by their unconscious knowledge of haunted locations (e.g. knowledge acquired earlier but now forgotten). However, due to the difficulty of assessing unconscious knowledge, and for ease of expression, we will use the phrase 'prior knowledge' throughout this article to refer to prior conscious knowledge.

Others have challenged the 'prior knowledge' hypothesis, noting that witnesses often claim to have been unaware of the reputation of a 'haunted' building prior to their experiences (see e.g. MacKenzie, 1982). This position has recently received empirical support from several studies conducted by Maher and her colleagues (for a review of these experiments, see Maher, 1999), using a quantitative technique pioneered by

Schmeidler (1966). In these experiments, mediums (individuals claiming to be sensitive to the presence of ghosts) were asked to walk through a 'haunted' building and mark floorplans to indicate where they felt a ghostly presence. These locations were then compared to the 'haunted' areas of the building (i.e. places in which witnesses had consistently reported ghostly phenomena). These experiments were not conducted in well-known 'haunted' buildings and none of the participants had any prior knowledge about the locations in which witnesses had reported 'ghostly' phenomena. Nevertheless, results from several studies demonstrate a significant relationship between the locations identified by the mediums and the 'haunted' areas. These findings suggest that many alleged hauntings may be the result of some people responding to some form of 'environmental cue' present in apparently 'haunted' areas (Lange *et al.*, 1996). Writers and researchers have suggested a huge range of factors to which people may be responding (for a review, see Houran, 1997). Some have suggested that these locations are actually haunted, and that people are responding to the presence of a discarnate spirit (e.g. Roberts, 1990). In contrast, others have suggested more mundane possibilities, including, for example, that these areas are simply rather cold and draughty (e.g. Nickell, 2001; Underwood, 1986). Others have speculated about the potential role played by rather more controversial physical factors, including, for example, low frequency sound waves (Tandy, 2000; Tandy & Lawrence, 1998), radioactivity (Radin & Roll, 1994) and local magnetic fields (Roll & Persinger, 2001).

A third part of Expt 1 thus examined the potential relationships among the 'haunted' areas, participants' reports of unusual experiences and magnetic fields. Measuring of the local magnetic field activity (i.e. all fluctuations between the range 0 to 3 kHz, whether of natural or artificial origins) was carried out because a relatively large amount of research has suggested a strong relationship between alleged hauntings and magnetic fields within this range. This work dates back to the mid 1980s, when Persinger (1985) speculated that changes in geomagnetic fields (created e.g. by tectonic stresses in the earth's crust) could stimulate the brain's temporal lobes and produce many of the subjective experiences associated with hauntings. Others have extended these ideas to account for physical manifestations including, for example, cold spots, electrical effects, popping sounds, etc. (see e.g. Houran & Lange, 1998). In a preliminary test of this theory, Gearhart and Persinger (1986) examined large case collections of alleged hauntings, and reported finding significant relationships between the time of onset of unusual phenomena and sudden increases in global geomagnetic activity (for a critique of this and related work, see Rutowski (1984) and Wilkinson & Gauld (1993)). More recent support has come from several on-site investigations of alleged hauntings that have reported unusual local magnetic activity (for an overview, see Roll & Persinger, 2001). Some of this work has noted that the effect seems to be associated with high levels of magnetic activity (Halgreen, Walter, Cherlow, & Cranall, 1978: Konig, Fraser, & Powell, 1981), whilst other researchers have related the effect more to variance in magnetic fields (see e.g. Persinger, 1985).

In Expt 1, the mean strength and variance of the magnetic field was measured in the Haunted Gallery and the Georgian Rooms. It was predicted that there would be significant differences between the mean field strength and variance in the 'haunted' and 'control' areas. It was also predicted that there would be a significant correlation between the number of unusual experiences reported by participants in each area, and the mean strength and variance of the magnetic field in those areas.

Method

Classifying 'haunted' and 'control' areas

IF had catalogued a large number of reports of unusual phenomena experienced by staff and visitors at Hampton Court Palace (Franklin, 1998). These reports dated from the end of the last century to the present day, and consisted of material from newspapers, magazines, books and IF's interviews with witnesses. Prior to the experiment, RW asked IF to identify where in the Haunted Gallery and the Georgian Rooms people had consistently reported unusual experiences. The palace supplied floorplans of both the Haunted Gallery and the Georgian Rooms. RW divided each of these floorplans into 24 equally sized areas and asked IF to mark the areas in which people had consistently reported unusual experiences. Areas marked by IF were classified as 'haunted' whilst unmarked areas were classified as 'controls'. IF marked seven areas in the Haunted Gallery and six areas in the Georgian Rooms. These floorplans were not seen by the investigators until all data collection had been completed. To avoid bias, RW, the assistant experimenters, who guided participants to the locations, and PS, who mapped the magnetic fields, were blind to the identity of these areas.

Questionnaires

In Questionnaire 1 participants rated the degree to which they knew where, in the Haunted Gallery or Georgian Rooms, people had experienced unusual phenomena in the past (definitely yes, probably yes, uncertain, probably no, definitely no).[1]

Questionnaire 2 asked participants to quietly walk around the Haunted Gallery or the Georgian Rooms, write a brief description of any unusual phenomena they experienced, indicate whether they believed that their experience(s) were due to a ghost (definitely yes, probably yes, uncertain, probably no, definitely no) and mark where they were standing when they had their experience(s) on a floorplan. The floorplan included in this questionnaire had not been divided into areas.

Procedure

Participants were self-selecting members of the public visiting Hampton Court Palace in late May/early June 2000. They had seen leaflets inviting their participation; thus they knew they were taking part in a scientific investigation. Participants took part in one of three daily sessions held over the course of 6 days. Each session involved a maximum of 40 people. Participants were first randomly split into two groups, according to where they had chosen to sit, with one half of the room forming Group 1 and the other half forming Group 2, in a counterbalanced order. All participants then completed Questionnaire 1, with Group 1 being asked about their prior knowledge concerning the Haunted Gallery, whilst Group 2 were asked about the Georgian Rooms. RW then gave a short talk about scientific research into ghosts. The talk was presented in an atmospheric setting, with lowered lighting. RW briefly described the historical tale of Catherine Howard, as outlined in the introduction, but without mentioning the location in which related haunting-type experiences had reportedly occurred. The talk also

[1] *Questionnaire 1 also contained other items including whether participants believed that ghosts exist, the frequency with which they had experienced 'ghostly' phenomena in the past, etc. The results of these items and related analyses are reported in Wiseman, Watt, Greening, Stevens, and O'Keeffe (2001). In summary, participants reporting prior belief in ghosts reported significantly more unusual phenomena during the experiment than disbelievers, and were significantly more likely to attribute the phenomena to ghosts.*

illustrated some of the apparatus that could be used in haunting investigations, such as heat-sensitive cameras and instruments sensitive to magnetic activity. Finally, RW outlined the purpose and methodology of the experiment.[2] Participants were then escorted by an assistant experimenter to either the Haunted Gallery (Group 1) or the Georgian Rooms (Group 2). Once at the location, the participants were free to walk around the location according to their individual preferences, and completed Questionnaire 2. Although participants were able to drop out of the experiment at any time without penalty, none did. Assistant experimenters were always on hand if needed by participants when they were in the test locations. Participants were also given RW's contact details in case they required further advice or information following the conclusion of the studies.

The first six sessions were pilot sessions, whose purpose was to check the practicability of the planned protocol, and to help identify areas for placement of measurement equipment. Data from the pilot sessions are not included in the analyses reported below.

Mapping the magnetic fields

Local magnetic fields were measured using two Mag-03MS100 3-axis sensors feeding into a laptop computer via a Mag-03DAM Data Acquisition Module (Bartington Instruments, Witney, Oxford). This system had a measuring range of 100 μT with a resolution of 0.1 nT, and recorded both static and dynamic components of the local field between 0 and 3 kHz. The system had a sampling rate of 1 Hz. Each sensor produced three streams of data, corresponding to the x, y and z axes of the local magnetic field, with a sampling rate of once per second. The three data streams were then combined to give the total field strength (using the formula sqrt [x(squared) + y(squared) + z(squared)]), and the mean field strength and variance was calculated from the resulting values.

Because the experimenters remained blind to IF's classification of 'haunted' and 'control' areas, it was necessary to find another way of selecting areas for placement of the instruments. Due to security and safety considerations, it was only possible to place the magnetic sensing equipment in 12 areas (six areas in the Georgian Rooms and six in the Haunted Gallery). These areas were agreed upon by RW and Hampton Court Palace administration on the basis of three criteria. First, to maximize the chances of detecting any anomalies in the magnetic fields, many of the areas chosen were those associated with a large number of reports of unusual experiences, derived from the pilot sessions; others were associated with a low number of reported unusual experiences. Secondly, to help minimize visitor disruption, the areas were not located in especially busy or narrow parts of the Haunted Gallery or the Georgian Rooms. Thirdly, to minimize the amount of time the equipment was in place, the areas were chosen such that they could be mapped in adjacent pairs.[3] All measurements were made by PS, who was blind to the

[2] *As an additional investigation of the effects of suggestion on reported ghostly experiences, during his introductory talk to participants RW made suggestions that one of the two locations was 'active' while the other was 'inactive' (in terms of recent frequency of reported ghostly experiences, but giving no specific suggestions as to where in each location experiences had been reported). To avoid systematic bias, these suggestions were made in a counterbalanced fashion. For sake of brevity, and because suggestion appeared to have little effect on reported experiences, this manipulation will receive no further attention in this article. More detail of the method and results of this manipulation can be found in the article by Wiseman* et al. *(2002).*

[3] *Once IF's classification of 'haunted' and 'control' areas was revealed at the end of the study, it transpired that the 12 areas chosen by RW consisted of six haunted and six control areas as identified by IF. The analyses for magnetic fields therefore refer to this 'sub-group' of six haunted and six control areas and not to the entire group of 13 haunted and 35 control areas.*

number of unusual experiences reported in each of the areas whilst setting-up and operating the magnetic field sensors. Magnetic data was recorded for thirty minutes in each area. Recording took place while tourists were visiting the area, but not during any experimental sessions. Hence the magnetic measurement procedure would not bias participants' reports.

Participants

There were 678 participants who each attended 1 of the 18 sessions. Some of the participants (131) were excluded as they did not complete all of the items on Questionnaire 1 and a further 85 were excluded for not completing all of the items on Questionnaire 2. The number of participants remaining was 462 (163 males, 299 females; mean age: 35.0, age range: 7 to 82, $SD = 16.3$). As the 18 groups of participants were assigned to one of the two locations, there was a total of 36 groups of participants.

Results

Participants reported a total of 431 unusual experiences: 189 (43.8%) of these experiences were reported in the Haunted Gallery and 242 (56.2%) in the Georgian Rooms; 215 (46.5%) participants reported at least one experience, and the mean number of experiences for participants reporting one or more experiences was 2.0 ($SD = 1.45$). Approximately two thirds of these experiences involved an unusual change in temperature. The remaining one third involved a mixture of phenomena including, for example, a feeling of dizziness, headaches, sickness, shortness of breath, some form of 'force', a foul odour, a sense of presence and intense emotional feelings. When asked whether their experiences were due to a ghost, 8 (3.72%) participants indicated 'Definitely yes', 22 (10.23%) 'Probably yes', 80 (37.21%) 'Uncertain', 87 (40.46%) 'Probably no' and 18 (8.37%) 'Definitely no'. It is difficult to assess the extent to which these experiences may have been elicited or dampened by the context of the pre-experiment talk. However, it is worth noting that both locations were well lit and relatively noisy and busy with tourists and were therefore less atmospheric than the context in which the talk was given. Given these circumstances, it was perhaps surprising that so many participants reported having experiences.

Participant grouping

Each of the 36 groups completed Questionnaire 2 whilst walking around either the Haunted Gallery or the Georgian Rooms. Individual responses to the questionnaire cannot therefore be considered statistically independent as they may have influenced, and been influenced by, other members of the group. For example, friends and family members were likely to have sat beside one another and therefore to have been assigned to the same group, so they may have interacted more with one another than strangers might. As a result, participants' responses to the questionnaire were combined within each of the 36 groups so the group is the unit of analysis (see Rosenthal & Rosnow, 1991).

Percentage of experiences reported in 'haunted' areas

The floorplans that had been divided into 24 areas were photocopied onto acetate and used to classify the location of each of the experiences reported by participants. This

classification was carried out by EG and CO, whilst blind to both the location of the 'haunted' and 'control' areas and the results of the magnetic field measurements. Given that there were seven 'haunted' areas in the Haunted Gallery and six in the Georgian Rooms, single mean *t* tests were used to compare the actual percentage of experiences reported in these areas with the chance baselines of 29.16% and 25% respectively. Both analyses found the percentage of experiences to be significantly greater than chance (see Table 1).

Table 1. The *df*, population means, *t* values (single group) and *p* values (two-tailed) comparing the percentage of experiences reported in the 'haunted' areas of the Haunted Gallery and the Georgian Rooms against chance

	% in haunted areas	Degree of freedom	Population mean	*t* test (single group)	*p* (two-tailed)
Haunted Gallery	38.83	16	29.16	2.954	.009
Georgian Rooms	46.24	17	25	3.494	.003

Prior knowledge

Each group's 'prior knowledge score' consisted of the mean of participants' responses to the question concerning the extent to which they knew where other people had reported unusual experiences in either the Haunted Gallery or the Georgian Rooms (coded on a 5-point scale from 1 (definitely yes) to 5 (definitely no)). Each group was then classified as having either 'High' or 'Low' levels of prior knowledge on the basis of a median split. This resulted in 18 groups being classified as 'High' (mean score = 3.89, $SD = .33$) and 18 groups as 'Low' (mean score = 4.51, $SD = .18$). There was a nonsignificant difference between the percentage of experiences reported in the 'haunted' areas by the 'High' and 'Low' levels of prior knowledge groups in either the Haunted Gallery ($t(15) = 1.66$, unpaired, $p = .12$, two-tailed) or the Georgian Rooms ($t(16) = -.14$, unpaired $p = .89$, two-tailed).

Magnetic fields

There was a nonsignificant difference in the mean magnetic field strength between the 'haunted' and 'control' areas (unpaired $t(10) = 1.55$, $p = .15$, two-tailed). However, there was a significant difference in the variance of the field between the two types of areas (unpaired $t(10) = 2.34$, $p = .04$, two-tailed), with the 'haunted' areas ($M = 12.71$, $SD = 12.10$) displaying a higher variance than 'control' areas ($M = 2.16$, $SD = 1.03$).

Spearman rank correlation coefficients were calculated between the number of experiences reported by each group within each of the 12 areas for which magnetic data was obtained, and mean strength and variance of the magnetic field in those areas.[4] One sample *t* tests were then used to examine whether the sample mean of these correlations differed significantly from zero. These analyses revealed a nonsignificant

[4] *There were three groups for which no experiences were reported in the 12 areas. As it was not possible to calculate a correlation in these cases, these three groups were not included in the analyses.*

relationship between the number of experiences reported and the mean field strength (1 sample $t(32) = .82, p = .42$, two-tailed). A significant relationship was found between the variance of the field and number of unusual experiences reported (1 sample $t(32) = 2.15, p = .04$, two-tailed).

Discussion

The experiment first examined whether participants would report a disproportionately large number of unusual experiences in the 'haunted' areas. These 'haunted' areas had been classified on the basis of prior reports. By chance, it was expected that approximately 29% of participants' unusual experiences would be reported in the 'haunted' areas of the Haunted Gallery, and 25% in the Georgian Rooms. However, groups of participants visiting both rooms reported significantly more unusual experiences in the 'haunted' areas within both locations. These findings strongly support the notion that people's unusual experiences are not evenly distributed across the locations, but instead concentrate in 'haunted' areas. In addition, the findings suggest that the areas in which people report their experiences are consistent across time. In short, these empirical findings validate several characteristics of spontaneous haunt experiences suggested by anecdotal reports.

Prior to entering either the Haunted Gallery or the Georgian Rooms, participants were asked to rate the degree to which they knew where people had reported unusual experiences in these locations in the past. The results showed that participants' level of prior knowledge was not significantly related to the percentage of experiences reported in the 'haunted' areas. These findings do not support the notion that the disproportionately large number of unusual experiences reported in 'haunted' areas is due to participants' prior conscious knowledge about the location.

Thirdly, the experiment examined the possibility that there were significant differences between the strength and variance of the magnetic fields between the 'haunted', and 'control', areas. Results suggested no significant differences in the mean strength of the magnetic field between the two types of areas. However, the variance of the local magnetic field was significantly greater in 'haunted' than 'control' areas, and there was a significant relationship between the magnetic variance and the mean number of unusual experiences reported by groups of participants. These results seem consistent with previous research suggesting a relationship between local magnetic field activity and haunt reports.

Experiment 2 (see below) built upon both the methodology and results of Expt 1. First, in Expt 1, areas within the Haunted Gallery and the Georgian Rooms were classified as either 'haunted' or 'control'. Experiment 2 provided a more fine-grained classification of areas by using a venue in which it was possible to rank order each of the areas from 'most' to 'least' 'haunted'. Secondly, in Expt 1, the nature of the venue resulted in participants having to walk around each of the locations in groups, and thus their data had to be analysed and interpreted at a group level. Unfortunately, this resulted in the study having low statistical power, and it is possible that the locations, having tourists as well as up to 20 participants walking around, were relatively noisy and therefore not conducive to haunt experiences. These issues were overcome in Experiment 2 by using a venue in which participants could visit areas on their own, and thus produce data that could be analysed and interpreted independently. Finally, Expt 2 measured a far greater number of environmental variables.

EXPERIMENT 2

The experiment took place in 10 of the South Bridge Vaults in Edinburgh. For the past few years, the company conducting guided tours through the underground vaults has maintained a collection of any unusual experiences reported by both guides and visitors. Prior to the experiment, RW asked Fran Hollinrake (FH), a senior tour guide, to review this database and rank order the vaults between 1 ('least haunted', i.e. smallest number of unusual experiences) and 10 ('most haunted', i.e. largest number of unusual experiences). This was referred to as the 'Haunted Order' of the vaults.

During the experiment, participants were asked to spend approximately 10 min in one of the vaults on their own, write down any unusual phenomena they experienced and rate the degree to which they believed that these experiences were due to a ghost. On the basis of the results obtained in Expt 1, it was predicted that there would be a significant correlation between the 'Haunted Order' and mean number of experiences reported in each vault. That is, it was predicted that the location of past haunt reports would be predictive of the location of haunt reports in the current study.

The experiment also investigated the potential relationship between participants' prior knowledge about the vaults and their reports of unusual phenomena. Prior to visiting the vaults, participants noted whether they knew where people had reported unusual experiences in the vaults in the past. Based on the results of Expt 1, it was predicted that the correlations between the 'Haunted Order', and the mean number of experiences reported in each vault, would be significant among participants who indicated no prior knowledge of the vaults.

The experiment also examined a wider range of environmental variables than Expt 1, including, the mean strength and variance of the local magnetic field, air temperature, air movement, the vaults' interior lighting levels, the lighting level directly outside the entrances to the vaults, the floorspaces of the vaults and their height. It was predicted that there would be significant correlations between these variables and both the 'Haunted Order', and the mean number of reported experiences in each vault.

Method

Questionnaires

Questionnaire 1 asked participants whether they had heard (e.g. from friends, the media, publications about the vaults) where in the vaults people have reported experiencing unusual phenomena (possible responses: yes, uncertain, no).[5]

Questionnaire 2 instructed participants to spend a few minutes in a vault and then report any phenomena that they experienced. They were asked to report all of their unusual experiences, no matter how faint, and to include all types of experiences (including e.g. unusual changes in temperature, smells, tastes, a sense of presence, etc.). The questionnaire contained four boxes, and participants were asked to briefly describe each of their experiences in one of the boxes. They were also asked to rate whether they thought that each of their experiences was due to a ghost (definitely yes, probably yes,

[5] *Other items on the questionnaire asked participants whether they believed in the existence of ghosts, whether they believed that they had previously experienced a ghost, etc. The findings will be reported in a separate article.*

uncertain, probably no, definitely no). If participants did not experience anything unusual then they were instructed to simply return the blank questionnaire.

Procedure

The experiment was carried out in April 2001. Participants were self-selecting members of the public who had seen the experiment listed in the programme of the Edinburgh International Science Festival. Participants took part in one of six daily sessions held over the course of 4 days. Each session involved a maximum of 10 people. The first part took place in a private function room close to the vaults. At the start of the experiment, RW handed out numbered clipboards randomly, which assigned a participant number to each person. RW briefly outlined the purpose and procedure of the study, and demonstrated the kinds of apparatus that could be used in scientific research into ghosts. RW then asked participants to complete Questionnaire 1. Participants were then taken as a group down to the vaults by FH, and then taken individually to a vault according to their randomly assigned participant number (i.e. participant number 1 went to vault 1). Note that RW was blind to the haunted order so he could not introduce bias by, say, assigning apparently suggestible participants to particular vaults. FH was not blind to the haunted order, but due to uneven flooring and low ceilings in parts of the vaults her presence was needed for safety insurance reasons, and she had very limited interactions with participants. Participants spent approximately 10 min in the vault and completed Questionnaire 2. During this time FH retired to a separate area of the vaults so she did not inadvertently influence participants' reports. Two assistant experimenters, who were blind to the haunted order, monitored participants while they completed their questionnaire and were available in case anyone had a query or a problem. Participants then returned their questionnaires to the assistant experimenters. Participants were able to drop out of the experiment at any time without penalty. Two did so. Participants were also given RW's contact details in case they required further advice or information following the conclusion of the studies.

Apparatus

Magnetic fields, air temperature and air movement

Local magnetic fields were measured using the same equipment as employed in Expt 1, but with an increased sampling rate of 4 Hz. Air temperature and air movement were measured with a Testo 445 multi-purpose datalogger connected to a Testo Hot Bulb probe (temperature range: −20 to +70C, movement: 0 to 10 m/s: accuracy, sampling at a rate of 0.5 Hz). Both the magnetic sensors and air temperature/movement probe were placed into one vault prior to each group's arrival. The participant in the vault was asked to remain a few feet from the equipment to prevent potential artifacts. The equipment logged data for 10 min. All measurements were made by PS, who was blind to the number of unusual experiences reported in each of the areas whilst setting up and operating the equipment. The magnetic sensor was sited at head height on a level part of the floor, at least 1 m away from the participant and on the opposite side of the room to any lighting circuits. When the participant arrived, PS started the recording and then left the vault.

Light readings and physical dimensions

The light levels within, and directly outside, each vault were measured using a Vital Technologies Corporation Tricorder. At the end of the experiment, RW recorded the light levels and physical dimensions of each vault. Light levels were recorded from the centre of each vault, and involved pointing the light meter towards each of the walls of the vault and taking an average of the readings obtained. The light level directly outside the vault was obtained by placing the light meter in the centre of the vault and pointing it towards the doorway of the vault.

Participants

The participants ($N = 218$) each attended one of the 24 sessions in groups of up to 10 (91 males, 127 females); mean age: 35.3 ($SD = 13.20$, age range: 11 to 77).

Results

Participants reported a total of 172 unusual experiences: 95 (43.58%) participants reported at least one experience, and the mean number of experiences for participants reporting one or more experiences was 1.81 ($SD = .94$). Again, the majority of these experiences involved an unusual change in temperature, but also included descriptions of apparitions, a strong sense of being watched, burning sensations, strange sounds, odd odours, etc. When asked to rate whether experiences were due to a ghost, 1 (.67%) experience was rated 'Definitely yes', 4 (2.67%) 'Probably yes', 58 (38.67%) 'Uncertain', 65 (43.33%) 'Probably no' and 22 (14.67%) 'Definitely no'.

Hypotheses

The correlation between the 'Haunted Order' and the mean number of unusual experiences reported in each vault, was significant ($N = 10$, rho $= .76$, $p = .02$, two-tailed).

Prior knowledge

Participants indicating 'yes' or 'uncertain' to the question regarding prior knowledge about where in the vaults people had experienced unusual phenomena in the past were then excluded from the data ($N = 31$). The correlation between the 'Haunted Order' and the mean number of unusual experiences reported by the remaining participants was highly significant ($N = 10$, rho $= .87$, $p = .009$, two-tailed).

Environmental variables

Table 2 contains the correlations between each of the environmental variables, and both the 'Haunted Order' and the mean number of experiences reported by participants. Overall, the magnetic field readings varied from 47,018-51,588 nT, *SD* from 4-32 nT. All of these measurements are within the natural fluctuation ranges and are not inherently anomalous. This is to be expected given that the vaults had no mains wiring other than a single, minimal lighting circuit.

Table 2. Spearman rank correlation coefficients (corrected for ties), and two-tailed *p* values (in parentheses), between each of the environmental variables, and both the 'Haunted Order' and mean number of unusual experiences reported by participants with no prior knowledge of the vaults. Statistically significant values are highlighted in bold

	Correlation with 'Haunted Order' ($N = 10$)	Correlation with mean number of unusual experiences ($N = 10$)
Magnetic mean	−.2	−.33
	(.55)	(.32)
Magnetic variance	.53	.39
	(.11)	(.24)
Air temperature	−.22	−.10
	(.50)	(.76)
Air velocity	.16	.43
	(.63)	(.19)
Interior light levels	−.50	−.26
	(.13)	(.43)
Exterior light levels	**.74**	**.84**
	(.03)	**(.01)**
Floorspace	**.73**	.58
	(.03)	(.08)
Height	**.65**	**.64**
	(.05)	**(.05)**

GENERAL DISCUSSION

The results of Expts 1 and 2 are highly consistent, with around 45% of participants in each experiment reporting at least one unusual experience. Some of these experiences were powerful for participants and were interpreted as being due to a ghost. Regardless of their interpretation by participants, all of these experiences are important and relevant to the question of the phenomenology of haunt experiences, as they give an indication of the incidence and nature of unusual experiences under controlled conditions in a potentially haunted location. In addition, it has been argued that the interpretation of such unusual experiences may be mediated by contextual variables (Lange *et al.*, 1996), such that the same experience may in one context be interpreted as ghostly, and in another context be interpreted as having a non-paranormal origin. Experiment 1 took place in a relatively well lit and busy location in which participants mingled as a group. In contrast, the setting for Experiment 2 was quiet, dank and dimly lit, and participants were alone while they rated their vault. While some aspects of the Hampton Court Palace setting may not have been conducive to ghostly experiences compared to the South Bridge Vaults, perhaps surprisingly a similar proportion of experiences was reported in each location. However, haunt experiences can and do occur in group settings and some authors have even suggested that group contagion effects may increase reports of haunt experiences (Lange & Houran, 1998, 1999). It is therefore possible that group contagion effects may have counteracted to some extent the less conducive aspects of the location in the Hampton Court Palace study.

The unusual experiences reported by participants in our two studies are comparable to many of the experiences that have been reported in the past in the two locations.

These past reports have contributed to the 'haunted' reputation of Hampton Court Palace and the South Bridge Vaults. Therefore, our findings can facilitate an understanding of these alleged hauntings. Our studies are perhaps less pertinent to highly documented cases in which a series of witnesses have reported seeing the same apparition over a long period of time (e.g. Gauld & Cornell, 1979; MacKenzie, 1982). However it has been noted (e.g. Beloff, 2001) that such cases are relatively rare. Also, the setting of our studies, in locations with haunted reputations, might not be directly comparable to those cases where unexpected experiences occur, for example when the experient had no prior conscious knowledge that the site might be haunted.

In Expt 1, participants reported a disproportionately large number of unusual experiences in 'haunted' areas. In Expt 2, there was a significant correlation between the 'Haunted Order' and the mean number of experiences reported in each vault. Together, these findings provide strong support for the notion that witnesses' reports of unusual experiences are not evenly distributed throughout the locations, but are instead concentrated in certain areas. In addition, they suggest that the locations in which their experiences are reported are highly consistent over time, as these are the areas in which most experiences have been reported in the past.

Both experiments also assessed the notion that this clustering of reports could have been due to participants having prior knowledge about where people have reported unusual experiences in the past. This idea has been proposed to account for many cases of alleged hauntings. However, the results from both experiments provided no support for this hypothesis. In Expt 1 there were no significant differences between the proportion of unusual experiences reported in the 'haunted' areas by groups of participants with 'high' and 'low' levels of prior knowledge. In the second experiment the correlation, between the 'Haunted Order' and the mean number of experiences reported in each vault, was significant among participants with no prior knowledge. Although we cannot rule out the possible effects of priming, expectation, and belief in the paranormal to account for people's reported experiences, these results strongly suggest that conscious 'prior knowledge' does not account for the clustering of experiences in certain locations within the two test sites. The finding that locations where experiences are reported is consistent over time, irrespective of prior knowledge, conceptually replicates previous fieldwork (e.g. Maher & Schmeidler, 1975; Schmeidler, 1966).

Thirdly, both experiments also examined whether the alleged haunting may be due, at least in part, to participants responding to environmental cues. In Expt 1, the variance of the local magnetic field in the 'haunted' areas was significantly greater than of the 'control' areas. In addition, the number of unusual experiences reported by participants was correlated with magnetic variance. This was not replicated in Expt 2, which found a nonsignificant positive correlation between magnetic variance and the haunted order. These results provide some support for the controversial theory that the presence of certain types of local magnetic fields may impact upon a range of psychological, psychophysiological and health-related variables (Korinevskaya, Kholodov, & Korinevskii, 1993; Voustianiouk & Kaufmann, 2000). A controlled laboratory study by Stevens (2001), for instance, showed psychological and physiological reactions to a changing magnetic field of comparable magnitude to those measured in our two experimental venues. Even subtle psychological and physiological changes occurring in a context that might suggest paranormal events (e.g. occurring to a person who believes in ghosts, occurring in a location with a haunted reputation) may lead to that person making a 'paranormal' attribution to what they might otherwise interpret as an

ambiguous stimulus. And it has been shown that experimentally applied weak magnetic fields can lead to more powerful and compelling experiences, such as a sensed presence, that are directly comparable to the kinds of experiences that are spontaneously reported (e.g. Persinger, 2001). Such findings suggest that magnetic fields, along with a range of other variables, together may account for some haunting experiences.

Results from Expt 2 also suggested that visual features of the environment may play a key role in causing people to report unusual experiences. The position of vaults in the 'Haunted Order' was positively correlated with the light level directly outside the vault, floorspace and height. In addition, the mean number of unusual experiences reported in the vaults was positively correlated with exterior light levels and height. These findings could be interpreted in several ways. For example, it is possible that these visual features might match the stereotype of a typically 'haunted' place held by participants, and thus induce mild psychosomatic and hallucinatory experiences. Alternatively, these features might directly cause unusual physical and psychological experiences. For example, participants walking from a relatively well-lit corridor into a much darker vault may cause them to experience the types of unusual phenomena associated with mild sensory deprivation (see e.g. Munro & Persinger, 1992; Tiller & Persinger, 1994). Likewise, especially large or high vaults may have caused participants to feel especially vulnerable and uneasy. Finally, these variables may covary with another factor (e.g. the production of unusual shadows) which are responsible for the reporting of unusual experiences. Future work should attempt to tease apart these competing interpretations of the phenomena by recording the number of unusual experiences reported by participants whilst systematically manipulating these factors (e.g. lighting levels and the variance of the local magnetic field). Multivariate modelling could be employed in future to understand the relative importance for haunt experiences of the variety of environmental and psychological factors that have been highlighted by investigations such as ours. As argued by Houran and Lange (1996), no single physical mechanism is likely to account for all cases of haunts.

In short, both of these experiments have yielded considerable insight into these two alleged 'hauntings'. Both experiments have demonstrated that the reputation of these locations is not based upon questionable eyewitness testimony, nor can the distribution of the experiences within the sites be explained by witnesses' prior knowledge. Instead, the data strongly support the notion that people consistently report unusual experiences in 'haunted' areas because of environmental factors, which may differ across locations. Further, our experiments have started to identify some of these factors, including the variance of local magnetic fields, size of location and lighting levels—stimuli of which witnesses may not be consciously aware. Taken together, these findings strongly suggest that these alleged hauntings do not represent evidence for 'ghostly' activity, but are instead the result of people responding—perhaps unwittingly—to 'normal' factors in their surroundings.

Acknowledgements

The authors would like to thank Ian Baker, Robert Chalmers, Dr Illya Eigenbrot, Ian Franklin, Christopher Gidlow, Ricky Glover, Fran Hollinrake, Dr James Houran, Dennis McGuinnes, Professor Robert Morris, Elizabeth Whiddett, Rachel Whitburn, Jeffrey Wiseman and our referees for their invaluable advice and assistance with these studies. We would also like to thank Bartington Instruments, Hampton Court Palace, the Edinburgh International Science Festival,

Mercat Tours, Land Infrared, L'Oriel Technology, Testo, the Perrott Warrick Fund, COPUS, and Philip Harris Education for supporting this work.

References

Auerbach, L. (1986). *ESP, hauntings and poltergeists.* New York: Warner Books.

Beloff, J. (2001). Foreword. In J. Houran & R. Lange (Eds), *Hauntings and poltergeists: Multidisciplinary perspectives.* Jefferson, NC: McFarland.

Franklin, I. (1998). *Hampton Court ghosts.* Unpublished manuscript.

Gallup (2001). *Americans' belief in psychic and paranormal phenomena is up over last decade* (http://www.gallup.com/poll/releases/pr010608.asp).

Gauld, A., & Cornell, A. D. (1979). *Poltergeists.* Boston, MA: Routledge & Kegan Paul.

Gearhart, L., & Persinger, M. A. (1986). Geophysical variables and behavior: XXXIII. Onset of historical poltergeist episodes with sudden increases in geomagnetic activity. *Perceptual and Motor Skills, 62,* 463-466.

Guiley, R. E. (1994). *The Guinness encyclopedia of ghosts and spirits.* London: Guinness Publishing.

Halgreen, E., Walter, R. D., Cherlow, D. D., & Cranall, P. H. (1978). Mental phenomena evoked by electrical stimulation of the human hippocampal formation and amygdala. *Brain, 101,* 83-117.

Henderson, J. (1999). *The town below the ground.* London: Mainstream.

Houran, J. (1997). Ambiguous origins and indications of 'poltergeists.' *Perceptual and Motor Skills, 84,* 339-344.

Houran, J., & Lange, R. (1996). Hauntings and poltergeist-like episodes as a confluence of conventional phenomena: A general hypothesis. *Perceptual and Motor Skills, 83,* 1307-1316.

Houran, J., & Lange, R. (1998). Rationale and application of a multi-energy sensor array in the investigation of haunting and poltergeist cases. *Journal of the Society for Psychical Research, 62,* 324-336.

Houran, J., & Lange, R. (Eds.) (2001). *Hauntings and poltergeists: Multidisciplinary perspectives.* Jefferson, NC: McFarland.

Houran, J., & Williams, C. (1998). Relation of tolerance of ambiguity to global and specific paranormal experience. *Psychological Reports, 83,* 807-818.

Irwin, H. J. (1999). *An introduction to parapsychology* (3rd ed.). Jefferson, NC: McFarland.

Konig, H., Fraser, J. T., & Powell, R. (1981). *Biological effects of environmental electromagnetism.* Berlin: Springer-Verlag.

Korinevskaya, I. V., Kholodov, Y. A., & Korinevskii, A. V. (1993). Effect of bilateral peripheral application of an alternating magnetic field on the EEG in humans. *Human Physiology, 19,* 213-217.

Lange, R., & Houran, J. (1997). Context-induced paranormal experiences: Support for Houran and Lange's model of haunting phenomena. *Perceptual and Motor Skills, 84,* 1455-1458.

Lange, R., & Houran, J. (1998). Delusions of the paranormal: A haunting question of perception. *Journal of Nervous and Mental Disease, 186,* 637-645.

Lange, R., & Houran, J. (1999). The role of fear in delusions of the paranormal. *Journal of Nervous and Mental Disease, 187,* 159-166.

Lange, R., Houran, J., Harte, T. M., & Havens, R. A. (1996). Contextual mediation of perceptions in hauntings and poltergeist-like experiences. *Perceptual and Motor Skills, 82,* 755-762.

Law, E. (1918). *The Haunted Gallery of Hampton Court.* London: Rees.

Lundberg, A. (1998). *The environment and mental health: A guide for clinicians.* Mahwah, NJ: Erlbaum.

MacKenzie, A. (1982). *Hauntings and apparitions.* London: Paladin Grafton.

Maher, M. C. (1999). Riding the waves in search of the particles: A modern study of ghosts and apparitions. *Journal of Parapsychology, 63,* 47-80.

Maher, M. C., & Schmeidler, G. R. (1975). Quantitative investigation of a recurrent apparition. *Journal of the American Society for Psychical Research, 69*, 341-351.

McCue, P. A. (2002). Theories of haunting: A critical overview. *Journal of the Society for Psychical Research, 66*, 1-21.

MORI (1998). Poll on the Paranormal. Commissioned by the *Sun* newspaper, February, 1998.

Munro, C., & Persinger, M. A. (1992). Relative right temporal-lobe theta activity correlates with Vingiano's hemispheric quotient and the 'sensed presence'. *Perceptual and Motor Skills, 75*, 899-903.

National Science Board (2000). *Science and engineering indicators.* Arlington, VA: National Science Federation.

Nickell, J. (2001). Phantoms, frauds, or fantasies? In J. Houran & R. Lange (Eds), *Hauntings and poltergeists: Multidisciplinary perspectives* (pp. 214-223). Jefferson, NC: McFarland.

Persinger, M. A. (1985). Geophysical variables and behaviour: XXII. The Tectonogenic Strain Continuum of Unusual Events. *Perceptual and Motor Skills, 60*, 59-65.

Persinger, M. A. (2001). The neuropsychiatry of paranormal experiences. *Journal of Neuropsychiatry and Clinical Neuroscience, 13*, 515-524.

Radin, D., & Roll, W. G. (1994). A radioactive ghost in a music hall. *Proceedings of the 37th Annual Conference of the Parapsychological Association*, 337-346.

Roberts, J. A. (1990). *Holy ghostbuster.* Dorset: Element.

Roll, W. G., & Persinger, M. A. (2001). Investigations of poltergeists and haunts: A review and interpretation. In J. Houran & R. Lange (Eds), *Hauntings and poltergeists: Multidisciplinary perspectives* (pp. 123-163). Jefferson, NC: McFarland.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

Rutkowski, C. A. (1984). Geophysical variables and human behaviour: XVI. Some criticisms. *Perceptual and Motor Skills, 58*, 840-842.

Schmeidler, G. R. (1966). Quantitative investigation of a 'haunted house'. *Journal of the American Society for Psychical Research, 60*, 137-149.

Stevens, P. (2001). Effects of 5s exposures to a 50 μT, 20 Hz magnetic field on skin conductance and ratings of affect and arousal. *Bioelectromagnetics, 22*, 219-223.

Tandy, V. (2000) Something in the cellar. *Journal of the Society for Psychical Research, 64*, 129-140.

Tandy, V., & Lawrence, T. R. (1998). The ghost in the machine. *Journal of the Society for Psychical Research, 62*, 360-364.

Thurley, S. (1996). *Hampton Court Palace.* London: Historic Royal Palaces Agency.

Tiller, S. G., & Persinger, M. A. (1994). Elevated incidence of a sensed presence and sexual arousal during partial sensory deprivation and sensitivity to hypnosis: Implications for hemisphericity and gender differences. *Perceptual and Motor Skills, 79*, 1527-1531.

Underwood, P. (1971). *The A-Z of British ghosts.* London: Chancellor Press.

Underwood, P. (1986). *The ghost hunter's guide.* Dorset: Blandford Press.

Voustianiouk, A., & Kaufmann, H. (2000). Magnetic fields and the central nervous system. *Clinical Neurophysiology, 111*, 1934-1935.

Wilson, A. J., Brogan, D., & Hollinrake, F. (1999). *Hidden and haunted: Underground Edinburgh.* Edinburgh: Mercat Tours.

Wilkinson, H. P., & Gauld, A. (1993). Geomagnetism and anomalous experiences, 1868-1980. *Proceedings of the Society for Psychical Research, 57*, 275-310.

Wiseman, R., Watt, C., Greening, E., Stevens, P., & O'Keeffe, C. (2002). An investigation into the alleged haunting of Hampton Court Palace. Psychological variables and magnetic fields. *Journal of Parapsychology, 66*, 387-408.

Received 18 September 2001; revised version received 13 February 2002

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[7]

THE 'BROTHER DOLI' CASE: INVESTIGATION OF APPARENT POLTERGEIST-TYPE MANIFESTATIONS IN NORTH WALES

by MICHAEL DANIELS

ABSTRACT

In February 1997 two Irish tourists allegedly witnessed a healing apparition of the Virgin Mary in a field near Mold, North Wales. Soon after, further apparitions and other phenomena began to be reported in the field and in the house of its owners. Since October 1998 a large number of unexplained stains and carvings of images and Welsh words, generally of a religious nature, have been discovered inside and outside the house. Welsh religious words were also found in e-mails sent from the house, in computer files and printed documents. Other phenomena have included noises, strange smells, temperature fluctuations, pools of water, electrical disturbances and object displacements. A number of photographs taken in and around the house also contained unexplained monk-like shapes and other seemingly anomalous images. An investigation of these various phenomena was undertaken. This focused on surveying the stains and carvings, on an examination of the photographic anomalies, and on an attempt to record manifestations using time-lapse surveillance equipment. Results are described, strengths and weaknesses of the evidence discussed, and possible interpretations considered. Although there are some highly strange features in this case, it is not possible to conclude with certainty whether the phenomena indicate genuine paranormal activity, whether they are the result of an elaborate hoax, or whether there is a mixture of genuine and fabricated incidents.

INTRODUCTION

As a result of articles in local and national newspapers and magazines, TV and radio broadcasts and coverage on several websites, the identity of the Gower family was already in the public domain prior to this investigation. Their real names are used in this report with their permission.

I was first contacted by e-mail on 11th October 2000 by Dr David Gower, a secondary school headteacher whose subject is science. In this e-mail, Dr Gower outlined the various phenomena that had been experienced at his home, most notably a series of stains and carvings of shapes and Welsh words, generally of a religious character, that were reported to have mysteriously appeared on walls and stones inside and outside the house. I visited the house for the first time on 12th November 2000 to view these phenomena and to talk about the case with Dr Gower and his artist wife Rose-Mary. Also living at home was the family's 15-year-old adopted son John-Paul, who has Down's Syndrome. A memorandum of agreement outlining the terms and conditions of an investigation was subsequently drawn up and signed by the Gowers and myself on 13th December 2000.

Between 12th November 2000 and 3rd March 2001, I made 14 visits to the house, totalling 54 hours on the premises. Throughout this period, regular e-mail contact was also maintained with the Gowers. As will be explained later, an ending to the phenomena was signalled to take place by 6th March 2001, and the investigation was wound down at that time. Further stains, carvings

Journal of the Society for Psychical Research [Vol. 66.4, No. 869

and other phenomena have since been reported, although these have been less frequent and are different in content. I made four further visits between 9th December 2001 and 12th May 2002 in order to check this report with the Gowers, to examine some of the more recent phenomena, and to discuss possibilities for future investigation.

Background to the Investigation

The following background account is based on interviews, conversations and e-mail communications with David and Rose-Mary Gower and their three daughters, and on newspaper reports and website articles. Because of John-Paul's condition, it has not been possible for me to question him directly on his experiences.

The first unusual events in the case were reported shortly after the Gowers first moved into their house in February 1997. Prior to this move the family had lived, since May 1995, in an adjacent bungalow.

On 7th March 1997 the lead article in the local weekly *Mold & Buckley Chronicle* reported that the newspaper had received a letter from an Irish couple, Michael and Concepta Dooley, who were on a walking holiday in North Wales. The letter described how, while walking along a lane (next to the Gowers' home), they stopped by a gate to admire the view. They then saw what seemed to be an apparition of 'Our Lady' at the top of the field. They were both overcome by a sense of awe, peace and serenity. Mr Dooley claims in the letter to have been cured of a painful frozen shoulder while his wife's cataract is also healing. Rose-Mary was interviewed by the newspaper and is reported to have remembered seeing the couple in the lane. She says they were in their late 50s or early 60s, seemed excited, but said nothing about their experience. Rose-Mary is also quoted as stating "I think it's all silly".

On 11th March 1997, a three-page hand-written letter addressed to "The Lady with the Labrador Dog" and signed "M & C Dooley" was forwarded from Wrexham to the Gowers. Dated March 9th, the letter gives the writers' address simply as "London", where they were visiting. It confirmed their sighting and adds that "we revisited the field behind the bungalow on Monday March 3 and while we did not see 'Our Lady' again we had a great sense of peace".

Figure 1. Face in the barn window (close-up right).

In March 1997, Rose-Mary took a series of 35 mm photographs of the house, garden, field (in which the vision had been seen) and the metal barn that is just beyond the field gate. One photograph appeared to show a large face in the far window of the barn (Figure 1). On 1st June 1997, this photograph appeared in the *Wales on Sunday* newspaper, under the headline: "Is this the face of Jesus' mother?" The article also reports that the original name of the bungalow (built c.1959) was "Santa Maria". On 17th May 1997, Rose-Mary says that she saw a white hazy figure standing alongside one of her daughters who had climbed to the top of the field with a friend. The two young women then ran down to say that they had sensed a presence. Also around this time, a number of visitors began turning up at the site to lay flowers and pray. On 8th June 1997, the *News of the World* ran an article: "Virgin Mary 'Cures' sick holiday makers". In November 1997, an account of the visions and cures also appeared in *The Paranormal Review* (Gower, 1997). Over the next year, several other newspaper articles appeared describing these religious visions.

In the summers of 1997 and 1998, the field became in Rose-Mary's words "a mini-Lourdes". In response to the number of visitors, the Gowers decided to set aside a small strip of land in front of the gate as a 'sacred site' (under the terms of the millennium Sacred Land Project established by the Archbishop of Canterbury and Prince Philip). A sign posted at the gate now welcomes those who wish to pray quietly or meditate outside the field.

The next series of phenomena began in August 1997 when Rose-Mary reports that a strange transformation occurred in the kitchen when she was clearing out a vase of dried blue flowers. Having placed the flowers on the floor while she stepped briefly outside, Rose-Mary returned to find that the blue petals had changed into a large group of dead and dying half-drowned wasps. On 31st January 1998, according to Rose-Mary, John-Paul saw a 'Blue Lady' by the barn door. Less than a week later, she found a small stone in the garden containing a discoloration and pockmarks that bore a resemblance to Mother Teresa of Calcutta. According to Rose-Mary, this was immediately recognised by John-Paul as the 'Blue Lady' he had seen a few days before. On 20th February 1998, Rose-Mary found another small stone in the shape of a coffin, containing a distinct cross-shaped indentation (both stones were examined and photographed by me but have since gone missing). On 1st September 1998, Rose-Mary reports that an oil-stain appeared on the driveway outside their house, which resembled an aeroplane with three vertical lines underneath the nose and tail (Figure 2). The following day, Swissair Flight 111 from New York to Geneva crashed off Halifax, Nova Scotia, killing all 229 people on board.

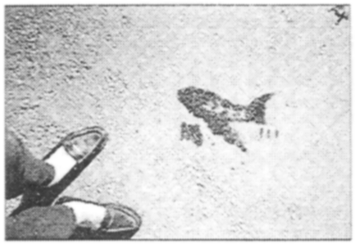

Figure 2. Aeroplane oil stain.

Journal of the Society for Psychical Research [Vol. 66.4, No. 869

In October 1998, the Gowers' youngest daughter, who was staying overnight in the house, awoke to see the figure of a young monk at the bottom of her bed. The apparition was initially experienced with a sense of great peacefulness, but it surprised her sufficiently to cause her to jump out of bed, scream, and turn on the bedroom light. A few days later, Rose-Mary reports that she saw through the kitchen window a fleeting apparition of the hooded head of what seemed to be a monk walking behind a hedge. This would have been impossible because there is a sudden drop from the base of the hedge down to the level of the river about 8 feet below.

In October or November 1998, a dark stain of a cross, about 12 cm high, appeared on one of the stones of the lounge fireplace. In January 1999, David, Rose-Mary and John-Paul returned home following a visit to the Midlands. They decided to move the furniture around and discovered a brown stain of a word, 'tangnefedd', about 60 cm in length, on the white-painted plaster of the lounge wall next to the fireplace (Figure 3, top). Presuming from the 'dd' at the end that the word was Welsh, they discovered that it meant 'peace' (in the religious sense). That month, about 20 further stains of Welsh words (generally of a religious nature), crosses and figural outlines that resembled a monk appeared on the same wall. Rose-Mary photographed them all and claims that sometimes words or images appeared in the photographs that were not apparent to the naked eye. In April 1999 a 30-cm tall monk-shaped stain appeared on a flat stone in the wall at the top of the stairs (Figure 4, left). Two days later, the Welsh word 'mynach' (monk) was found very neatly carved down the centre of the stained stone (Figure 4, centre). This rather beautiful stone is now generally referred to as the 'Monk Stone'.

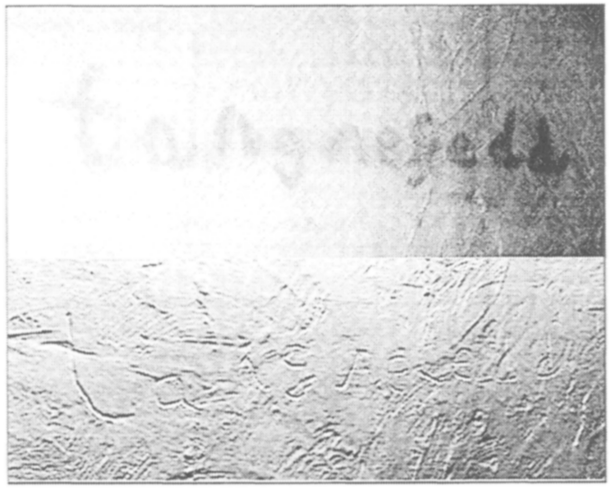

Figure 3. The original 'tangnefedd' stain and carving.

In the spring of 1999, Rose-Mary repainted the lounge walls using white emulsion, though carefully preserving the original 'tangnefedd' stain. Having

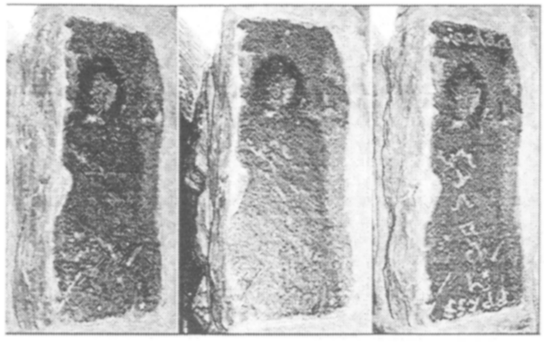

Figure 4. The 'Monk Stone', showing development of carvings (left to right).

completed the painting about 7 pm, she noticed that stains had reappeared on the wall, but now as different words in different places. In May 1999 David, Rose-Mary and John-Paul returned from a visit to Harrogate to discover that the 'tangnefedd' stain had disappeared, to be replaced by the same word now *carved* into the plaster of the wall (Figure 3, bottom—note the identical surface irregularities in both photographs). This was followed by the appearance of other plaster carvings, including crosses, monk-shaped outlines, and various other Welsh words of a religious nature such as 'gobaith' (hope) and 'cariad' (love). What was strange about these carvings was that the paintwork showed no evidence of having been scraped away, but continued apparently undisturbed inside the indentations. Also it is reported that no dust or debris was ever found in the vicinity of the carvings. Over the next several months, further carvings and stains appeared on the lounge walls and fireplace. Since this time, similar stains and carvings have appeared in other areas of the house, on outside walls, on stones in the garden, on the trunk of a holly tree, and an outside dog kennel.

A name was needed to enable the family to refer to whatever was responsible for these phenomena. Because of the religious nature of the words and images, the apparitions of the monk experienced by Rose-Mary and by her daughter, and the high frequency of monk-shaped stains and carvings, a monk seemed to be the most obvious candidate. The name 'Brother Adolphus' ('Brother Dolly' for short) was coined. This was a name that, as a child, Rose-Mary had noticed appearing every year in the *In Memoriam* column of a newspaper. Sometime later a message from the monk was found written on a notepad and signed 'Doli'. The family became accustomed to talking about Brother Doli as if he were an actual presence in the house, although David in particular has done this rather with tongue in cheek.

In August 1999, the phenomena at the Gowers' home were examined on HTV Wales' *Weird Wales* series. The case has since featured on several other TV and radio programmes, including ITV's *This Morning*, BBC TV's *Kilroy*, and John Peel's *Home Truths* series on BBC Radio 4 (twice).

Between 14th and 22nd October 1999, 39 e-mail messages were sent out to fourteen people in the Gowers' e-mail address book. These messages contained

Journal of the Society for Psychical Research [Vol. 66.4, No. 869

Welsh religious words or strings (e.g. 'mynach', or 'tangnefeddcariadmynach'). With the exception of those sent on the first day, these messages were not found in the Sent folder. The times recorded on the messages also indicated that several had been posted in the middle of the night. Since that time, e-mail messages sent by the Gowers have occasionally contained Welsh word intrusions, including one ('tangnefedd') in the first e-mail I received from Dr Gower. Other reported computer-related phenomena include deletion of files and the frequent appearance of Welsh religious words or religious images as intrusions on screen or in printouts of documents. Sometimes lettering is inverted or printed sideways.

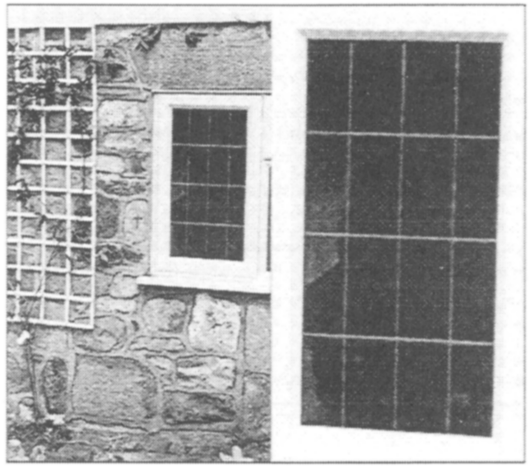

Figure 5. The monk in the lounge window (close-up right).

In March 2000, Rose-Mary took a 35 mm photograph of a stained cross that had appeared on the stonework outside the lounge window. When the print was examined, it seemed to show a dark outline, possibly resembling the shape of a monk, inside the house (Figure 5). Also a vase that normally appears in the window was missing in the photograph. Around the same time Rose-Mary also took a photograph of a carved wooden bench that stands at the top of the stairs. This was a family heirloom that had been given to the Gowers a few months earlier by Rose-Mary's mother, who had first acquired it in Guernsey around 1952. The family believed that the bench came from a monastery and it has always been called the 'monk's bench'. The photograph shows the apparently anomalous images of a white outline of a hooded figure and a crucifix on the backboard of the bench (Figure 6).

In the spring of 2000, a monk-shaped outline, about 1.5 m in height, appeared cut out of the lawn, as if with a spade. The word 'mynach' was also cut in the lawn, in letters about 30 cm high. In July 2000, the word 'iachâd' (healing) was found neatly carved across the top of the 'Monk Stone' at the top of the stairs. Two weeks later, the word 'ffydd' (faith) appeared at the bottom of this stone, again very neatly carved (Figure 4, right).

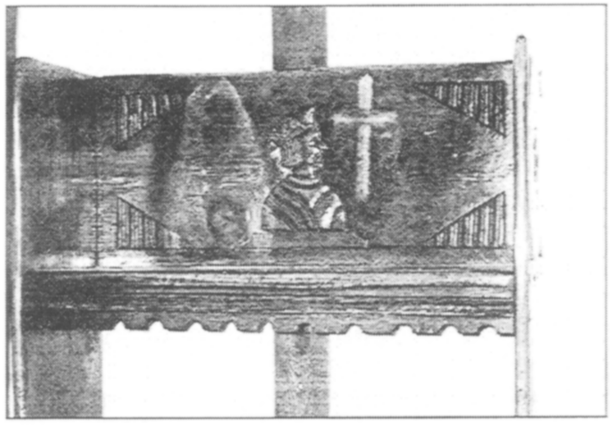

Figure 6. Monk's bench with monk and cross.

In September 2000, Rose-Mary took a 35 mm photograph of one of her digital artworks, which had been hung on a kitchen wall directly underneath a decorative mirror. The photograph showed a dramatic outline of a faceless white figure apparently reflected in the mirror (Figure 7, left).

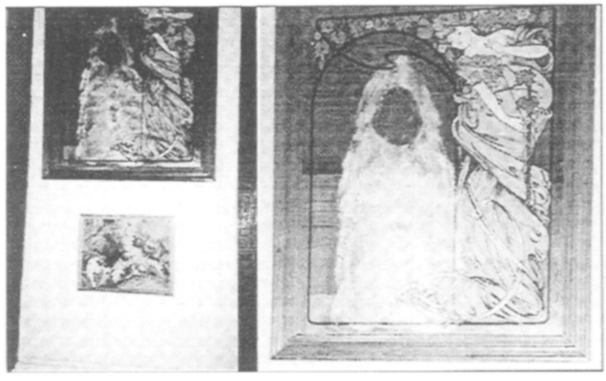

Figure 7. Monk in the mirror. Scan of original print (left) and scan of original negative (right).

In addition to the range of phenomena described above, which comprise apparitions, stains, carvings, photographic effects and computer-related anomalies, the Gowers also report:–

- **Noises** — including footsteps on the stairs and landing, latches being lifted, bangs and crashes, the sound of a baby crying in the kitchen, and indistinct voices.
- **Smells**—especially of incense and candle wax.

Journal of the Society for Psychical Research [Vol. 66.4, No. 869

- **Pressures**—feeling a pressure on the bottom of the bed, as if someone had sat down on it (experienced by Rose-Mary and a daughter).
- **Temperature changes**—sudden unexplained drops in temperature in the lounge.
- **Pools of water**—unexplained pools of water appearing on the floor inside the house, or on chairs and a bed.
- **Strange animal behaviour**—the cat often rushes around the house as if someone were chasing him.
- **Displaced objects**—personal items such as spectacles and slippers going missing, to be found later in unusual locations around the house.
- **Other graffiti**—visitors to the house have found Welsh words engraved on their personal belongings (e.g. purse, handbag and a laptop computer).
- **Electrical and telephonic anomalies**—clocks changing time or going slow, light bulbs flickering or flashing when switched off, silent phone calls made from the house, often recorded on recipients' answering machines.

THE INVESTIGATION

The House

The Gowers' home is situated in a very attractive and peaceful rural area, a few miles from the market town of Mold, Flintshire. It is a converted, extended and modernised period dwelling. Some of the original stonework is retained. The lounge is located in the oldest part of the house and features a beamed ceiling and a large reconstructed stone fireplace. Little is known about the history of the property, although the Gowers believe the original building was possibly a miner's cottage. A stone built into the front elevation has the date "1610", but this seems to be modern. Outside there is a garage and workshop. The narrow, fast-flowing River Terrig runs through the garden close to the house. Situated further up a lane and not directly adjoining the property is a steep field that belongs to the Gowers where the visions were experienced (approx. 2.5 acres). Just over the gate inside the field is a large metal barn where the face in the window was photographed. There are four neighbouring dwellings.

The Gowers

The Gower family is not Welsh, or Welsh speaking. David was born in 1947 and was brought up in Middlesex. There was a strict Baptist influence in his family, but he no longer sees himself as religious. He has a BEd in Chemistry and Divinity, MEd, PhD in Chemical Education, and MBA in Education Management (qualifications checked by me). He describes his attitude to the phenomena as scientific and sceptical, but he also finds the case entertaining and intriguing. His early suspicion was that his grown-up children were playing a hoax, but he now tends to discount this possibility. He admits to having witnessed strange noises in the house (e.g. footsteps, crying baby, indistinct voices), but has never experienced a visual apparition. His work means that he is typically away from the house for much of the day.

Rose-Mary was born in 1950. She reports that her family experienced poltergeist-type phenomena at their home in Guernsey from before the time she was born, throughout her childhood, and after she had left home. When she was about eight years old, Rose-Mary and her two younger sisters named the poltergeist 'Kelly'. 'Kelly' used to steal things and cause bangs and crashes. On one occasion, a mirror was smashed. Rose-Mary reports that she was frightened by these phenomena, in contrast to the present events, which seem harmless or benign. Rose-Mary's family was strictly Pentecostal. She first met David when she was 15, as a result of contact between the families. Rose-Mary is no longer affiliated to any particular religious group. Rose-Mary is typically in the house most of the day, where she has home-educated John-Paul. Since January 2000, Rose-Mary has discovered a surprising talent for digital art and for poetry. She has had exhibitions of her art and regularly sells her work, which is often commissioned. Rose-Mary attributes the sudden flowering of her artistic talent to the Brother Doli phenomena and, as a result, uses the name "Y Mynach" for her art business. Rose-Mary describes her attitude to these phenomena as one of interest and her view is that something very strange is happening at the house. She also attributes human qualities and personality to 'Brother Doli', whom she claims to have 'seen', fleetingly but distinctly, on five occasions including once, in the corner of the kitchen, when she and I were in conversation. Since taking the photograph of the face in the barn window, Rose-Mary will now often finish off a roll of film by taking photographs in and around the house.

The Gowers have three daughters (born 1970, 1974 and 1976) and two adopted sons (one born 1973 and John-Paul born 1985). The daughters and elder son left home between 1989 and 1997. The daughters remain regular visitors to the house, but the elder son has made few visits since he left home in 1994. I have briefly met and spoken with the three daughters. All three are interested in and intrigued by the phenomena and claim to have heard footsteps in the house. They also confirm the apparitional experiences previously described.

John-Paul was adopted in 1986, when he was 13 months old. He has quite severe Down's Syndrome. He can read, copy words and, with assistance, can use a computer to write letters. He is a quiet and gentle young person, and seemed to treat 'Brother Doli' in a very matter-of-fact way. When asked where Brother Doli was, John-Paul would typically claim to be able to see the monk, who, he stated, lived in a corner in his bedroom (where the daughter first saw the apparition of a young monk). I am told that he also sometimes reported on Brother Doli's moods and could be quite convincing when describing the monk's activities even though he is not noted for having a good imagination. John-Paul typically kept himself out of the way during my visits and seemed to show little understanding of or interest in the investigation. He amuses himself by watching videos or playing Nintendo. He is sometimes left alone in the house for an hour or two.

The Dooleys

Evidence from Michael and Concepta Dooley is currently limited to the testimony of Rose-Mary, who claims to have met them in the lane in late

Journal of the Society for Psychical Research [Vol. 66.4, No. 869

February 1997, together with information in the two letters apparently sent by them to the *Chronicle* and to "The Lady with the Labrador Dog". Unfortunately neither letter included a return address although the *Chronicle* letter states that they live in a village near Dublin. No further contact with the Dooleys has been forthcoming and my attempts to identify them from the Dublin telephone directory have not met with success. Of course, the Dooleys may have moved away or used an alias in their letters (the lack of a return address indicates that they did not wish to be contacted). The possibility remains, however, that the Dooley letters are a hoax perpetrated by the couple themselves, or by someone claiming to be them.

The Stains and Carvings

I surveyed the physical manifestations at the house during visits for this purpose on 13th, 14th & 19th December 2000, and on 7th & 14th January 2001. Records were made of all stains and carvings, including their location and approximate measurements. Stickers were used for the purpose of marking and identification. I photographed the majority of stains and carvings using Kodak Gold 800 film with a Canon EOS 1000F 35 mm SLR camera and Canon Zoom Lens EF 35–105mm 1:4.5–5.6. Rubbings were also taken of several prominent carvings in order to provide a more exact record of size, shape and surface details. Two full days (13th & 14th December) were needed to record all the phenomena then apparent. The visits on 19th December, 7th January and 14th January recorded stains and carvings that had appeared since the previous visits, as well as some that were revealed behind furniture not moved earlier because of Christmas preparations and decorations.

Table 1

Summary of Recorded Phenomena (13th December 2000—14th January 2001)

Type	**N**
Stain	148
Carving	57
Carving + stain	11
Burnt carving	4
Raised plaster	7
Other	7

Recorded	**N**
13th Dec 2000	72
14th Dec 2000	70
19th Dec 2000	32
7th Jan 2001	38
14th Jan 2001	22

Location (Floor) [1]	**N**
Lounge total (G)	134
(Lounge fireplace)	(69)
Kitchen (G)	49
Study (G)	7
WC (G)	1
Stairs	12
Landing (F)	4
Bathroom (F)	4
Bedroom 1 (F)	3
Bedroom 2 (F)	0
Bedroom 3 (F)	0
Bedroom 4 (F)	0
Outside	20

Surface	**N**
Plaster	109
Stone	99
Wood	15
Brick	4
Other	7

Content	**N**
Words	77
'Monk'	52
Cross	50
'Smiley face'	10
Monk with cross	6
Chi-Rho symbol	8
Others	31

Total N in each group = 234

[1] G = Ground floor, F = First floor

A total of 234 stains, carvings and other features were recorded during these visits. Additionally there were many features that were not included in the survey because they were too faded to identify clearly, or because they might have been natural artefacts. Table 1 summarises the phenomena recorded during the survey. To give an indication of the nature and extent of these phenomena, Figure 8 shows a view of the fireplace taken on 14th December 2000 (the fireplace has never been used by the Gowers). The small rectangles in Figure 8 are stickers identifying stains and carvings.

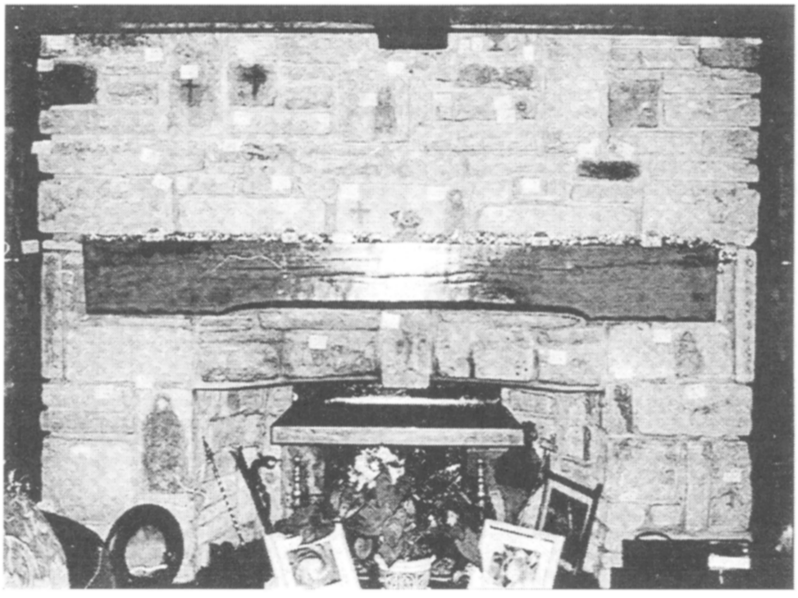

Figure 8. The lounge fireplace (at 14th December 2000).

The great majority of stains and carvings appeared in the older (original) part of the house, especially the lounge and the wall by the stairs (which was originally an external wall). Few were found upstairs. There seemed to be particular 'hot spots' in the house, most notably the fireplace, the area directly beneath the stairs, and a wall next to the kitchen door that leads to the patio.

Stains on painted plaster were generally coffee-coloured, although five plaster stains that appeared in the New Year of 2001 were a pale blue. Stains on stonework were usually mid-brown or rust-coloured, although some were almost black. Stains on wood were generally silver-grey. Size of stains varied from a tiny cross at 1 x 1 cm to two large monk-shapes (33 x 70 cm, 30 x 81 cm), and a large word, 'hirymaros' (long-suffering) at 98 x 13 cm (all measurements are width x height). There was also a large monk shape (48 x 108 cm) on a textured outhouse wall. This particular image was unusual in that it was unclear whether it was a stain or a shadow produced by unevenness in the wall's textured surface.

Journal of the Society for Psychical Research [Vol. 66.4, No. 869

Stains were found at various positions and heights, from floor to ceiling (including two on the carpet). One is reported to have appeared on the outhouse chimney, which is accessible only with difficulty. Following a 'throwaway' comment I made one day that it was a shame that the phenomena didn't appear in inaccessible places, stains soon started to appear behind radiators and were later discovered behind heavy furniture.

The family reports that the appearance of the stains often varied over time. The stains may fade or increase in vividness, may change in colour or size, and may disappear altogether. New stains may suddenly be noticed that did not seem to be there moments before. I can confirm the apparent fading, intensifying and disappearance of stains on both plaster and stone surfaces. It should be noted, however, that changes in lighting inside the house often produce noticeable variations in the appearance of the stains. New stains also appeared during the course of my survey, including one dramatic stain of a monk shape on the fireplace that was immediately noticed on the morning of 14th December 2001 as not having been there the evening before (confirmed by photographic evidence).

From casual inspection, the stains appeared to be produced by some agent that had soaked into the plaster, stone, brick or wood, rather than being a surface film. The stains could not be rubbed or washed off and are reported not to respond to bleaching. The chemistry of the stains was not analysed during my initial investigation and most have since faded. In recent months a few new stains have appeared. It is planned to undertake detailed chemical analysis on these, and on the few original stains that remain, for a future report.

'Carvings' in plaster comprised incisions that were generally about 2 mm or less deep and about 5 mm or less wide. Some carvings were quite crude, but others were precisely executed. They varied in size from a small cross at 2 x 3 cm to a large 'tangnefedd' at 62 x 10 cm (Figure 3, bottom) and occurred on the walls at different heights. In all cases the paintwork continued inside the indentations. To test how easy it would be to make such carvings in plaster, I experimented at my own home (an old Welsh cottage that has similarly plastered stone walls). My experiments showed that plaster was a surprisingly good medium to work and I was able to produce neat carvings easily and quickly using a bradawl or small screwdriver.

The question of whether the carvings have been overpainted is crucial in assessing whether they may be a hoax. In most cases, although there was no obvious indication of overpainting (i.e. the paintwork around and inside the carving did not appear to be new), there were some indications of a difference in texture around some of the carvings. For example, the area around the carved surface often appeared smoother. In a few cases, there was evidence of brush-strokes around the carving that differed in direction and fineness from those found on the surrounding surface. An example of this is shown in Figure 9.

A few carvings appeared to 'fade' during the course of my investigation (i.e. to become shallower), possibly also indicated in Figure 9. This was first noted following another 'throwaway' comment that disappearing *carvings* (as opposed to stains) would be very impressive. In one case, a previously very distinct carving of a monk had disappeared (could not be located) on my visit of 14th January 2001, although it later re-emerged, though less prominently than

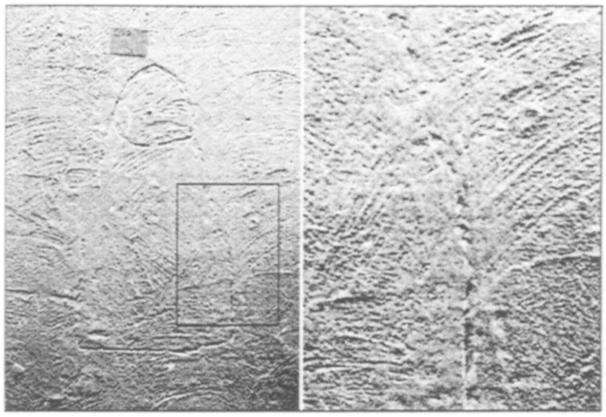

Figure 9. Carving of 'monk' (detail right) showing possible evidence of overpainting.

before. There seemed to be no obvious indication that the carving had been in-filled. However, on my latest visit (12th May 2002) there was evidence that a very crude attempt had recently been made to in-fill, skim and paint over the Figure 9 carving. Also on this day another fresh-looking plaster carving was discovered in the kitchen. There was a faint but noticeable smell of paint around the carving that was confirmed by members of the family.

Another type of plaster 'carving' appeared very slightly *raised* from the surrounding surface. The first of these was a monk shape that I noticed on 14th December 2000 (again following an earlier comment that raised surfaces would be more evidential than indentations). This particular 'monk' gave the appearance of a slight swelling in the plaster. On 14th January 2001, six further raised plaster features had appeared in the lounge. Unlike the first raised carving, however, these latest features seemed to have been built up using some kind of plaster or filler material. Generally they looked crude and unconvincing.

Carvings in stone varied in size from a 'smiley face' at 2 x 2 cm to the date '1778' at 23 x 4 cm. They were generally found in easily accessible locations at heights that would be relatively comfortable to work at. Many of the carvings were quite crudely executed and looked as though they could have been scraped away with a pointed tool. A few 'carvings' (e.g. '1778' and a 'face' on the fire-place keystone) appeared to be partly formed from natural edges on the stones.

The most interesting and significant carvings are undoubtedly the words on the 'Monk Stone' at the top of the stairs (Figure 4), which appeared to be executed with considerable skill and precision. Here the edges of the words were straight and crisp and the carvings made to an even, flat depth (about 1mm). They did not look fresh, although there were variations in coloration that matched the order in which they appeared. Thus 'mynach' (the earliest of the three words) was darker and older looking than 'iachâd' (which came second) while 'fydd' (the last to be discovered) was the lightest and freshest looking. The 'Monk Stone' itself is very unusual. The carved and stained

Journal of the Society for Psychical Research [Vol. 66.4, No. 869

surface is almost completely flat and appears to be faced with a thin layer of hard, darker stone. The family is particularly intrigued by this stone, which has become an interesting talking point. Rose-Mary polishes it regularly.

Carvings in wood were crudely executed and looked as though they could have been made with a chisel, knife, scissors or other scraping implement. They varied in size from a tiny cross at 0.4 x 0.3 cm, which could simply have been a pencil mark, to an 8 x 18 cm 'monk' cut into the trunk of a holly tree. They were generally found in accessible locations and at comfortable heights for working. One or two looked freshly made, although others appeared quite old. Some were varnished inside, matching the surrounding wood. Four examples of woodcarvings looked as though they had been burnt with a pyrography tool (the family possesses such a tool).

The lawn cuttings are another very unusual feature of this case. Two had been reported before my visits. On 14th December 2000 I discovered three new ones on the raised lawn. Fallen leaves that covered them indicated that they might have been there for some time. The first was a bare patch of lawn a few centimetres in depth, in the shape of a 'monk' (approx. 66 x 100 cm). Another was an edged outline of a cross shape (approx. 89 x 140 cm). Finally there was a layer of turf, about 4 cm thick, in the shape of a 'monk' (approx. 54 x 100 cm).

Table 2

Glossary of Words (Survey Results at 14th January 2001)

Word	Translation	N	Word	Translation	N
angel	angel	3	Jane	Jane	1
cariad	love	1	Jones	Jones	2
cerdd	poem, music	1	llanwenydd [1]	joy, gladness [1]	1
ci	dog	1	llun	form, image	1
croeso	welcome	8	maddeuant	forgiveness	1
daioni	goodness	2	Mari	Mary	2
dedwyddwch	happiness	1	MCDXVI [2]	1416 [2]	2
dysgu	learn, teach	1	mynach	monk	18
ffydd	faith	2	Nadolig	Christmas	1
gobaith	hope	1	perthyn	belong	1
goleuni	light	1	St Teresa	St Teresa	1
gorffwys	rest, repose	1	tangnefedd	peace	5
gweddi	prayer	4	teulu	family	5
gwyn	white, blessed	1	Treffynnon	Holywell	1
hirymaros	long-suffering	1	ysbryd	spirit, ghost	2
iachâd	healing	4	15, 1778, 2000 [3]	15, 1778, 2000 [3]	3

[1] This would appear to be a spelling mistake for the intended 'llawenydd' (joy, gladness).

[2] The date 'MCDXVI' is unusual in that the letters C and D have appeared as mirror images. The other four letters are, of course, laterally symmetrical. If viewed in a mirror, therefore, it reads IVXDCM.

[3] The number 2000 is reported to have appeared close to the millennium.

Although these latter measurements are not a precise match for those of the 'monk-shaped' bare patch, it seems likely that this layer was that removed from this patch of lawn. The appearance of these lawn features gave the impression that they had been carefully cut with a spade.

The Words and Numbers

The words and numbers recorded in the survey are shown in Table 2. Other words that have appeared since January 1998 are shown in Table 3 (these have been reported by the Gowers on webpages, in interviews or e-mails, or are shown in earlier photographs taken by Rose-Mary, or have been observed by me since the survey was undertaken).

Table 3

Additional Words (January 1998–May 2002)

Word	Translation	Word	Translation
Amwythig	Shrewsbury	gwynfa	paradise
anabl	disabled	Karol W [3]	Karol Wojtyla? [3]
arogidarthu [1]	burn incense	magwraeth	nourishment, nurture
bendith	blessing	mynachlog	monastery
cannwyll	candle	paderau	rosary
edifar	penitent, sorry	pererindod	pilgrimage
ffyddiondeb [2]	faithfulness	plentyn	infant, child
golau	light	rhinwedd	virtue
gwyddonydd	scientist	Sanct Mari	St. Mary
		Tom	Tom

[1] Correct spelling = arogldarthu.

[2] Correct spelling = ffyddlondeb.

[3] The family interprets this as a reference to Pope John Paul II (Karol Wojtyla)

On several occasions words appeared seemingly in response to current events. For example the number '2000' is reported to have appeared at the time of the millennium, 'Nadolig' (Christmas) shortly before Christmas, 'gwyddonydd' (scientist) during a visit of a scientist friend, and 'ci' (dog) the day after the death of the family dog.

The stained and carved words are written in a simple, but very distinctive and consistent, script that was even used, I am told, for the word 'mynach' that had appeared cut in the lawn. The letter 'n', for example, resembles a witch's hat. Lettering is always unjoined. All words (with the exception of 'Jane', 'Jones', 'St Teresa', and 'Karol') appear in the *Collins Gem Welsh Dictionary* (purchased by the family following the initial appearance of Welsh word stains). However, 'tangnefedd', which was the first word reported, appears in the dictionary only under its English translation (peace).

There are no examples of sentences or other syntactical constructions. In the case of the e-mail phenomena, Welsh words have been typically strung together in a seemingly haphazard fashion.

Journal of the Society for Psychical Research [Vol. 66.4, No. 869

The examples of what appear to be clear spelling mistakes are of particular interest (these errors have been confirmed by Dr Branwen Jarvis, Head of the Welsh Department at University of Wales Bangor, in a personal communication). In two cases, 'arogidarthu' (see Figure 10) and 'ffyddiondeb', the letter 'l' has been incorrectly replaced by 'i'. In other words, the actual Welsh words are 'arogldarthu' (to burn incense) and 'ffyddlondeb' (faithfulness). Because of the clear difference in pronunciation involved, this would not seem to be the kind of spelling mistake that a Welsh-speaking person could easily make. It should be noted that the font used in the *Collins Gem Welsh Dictionary* is small and the letters 'i' and 'l' might easily be confused by someone with poorish eyesight who was unfamiliar with the language.

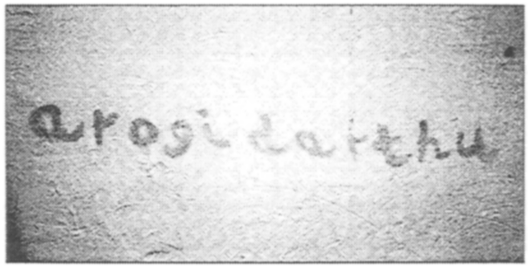

Figure 10. An apparent spelling mistake.

Another significant error occurs with 'llanwenydd', which is not a Welsh word. However 'llawenydd' means 'joy' or 'gladness' and this would appear to be the word intended. Given the frequency of Welsh place names beginning with 'llan' (church or village) it is very possible that a person aware of this but generally unfamiliar with Welsh could have mistakenly read the word from the dictionary. There is no such place as Llanwenydd in Wales.

The words 'Jane', 'Jones', the number 15, and the date 1778, which were found in close proximity on the fireplace, directly match words on a loose tombstone that leans against the outside front wall of the house. The origins of the tombstone are unknown, but its presence at the house predates the Gowers' occupancy (as shown by the original estate agents' photograph). The tombstone reads "Jane Jones / AGED 15 / 1778".

The date MCDXVI (1416), which appeared during the time of my investigation, is of especial interest when considered alongside some of the original word stains photographed by Rose-Mary. One photograph shows an arrow pointing between the words 'Amwthig' (Shrewsbury) and 'Treffynnon' (Holywell). Under these words is the word 'pererindod' (pilgrimage). This seems to refer to pilgrimages associated with the cult of St Winifred, which was especially prominent in the 12th–15th centuries (although pilgrimages to St Winifred's Well at Holywell continue to this day). According to legend, in the 7th century, Winifred (Gwenfrewi) was decapitated by Prince Caradoc after she resisted his attempts at rape. Winifred was brought back to life by the prayers of her uncle, St Beuno, while Caradoc was struck dead. Where Winifred's head fell, a spring broke forth (at Holywell). Winifred spent the

remainder of her life as a nun at Gwytherin in nearby Denbighshire. On her death (c. 680) Gwytherin and Holywell became important places of pilgrimage. In 1138, Winifred's remains were removed from Gwytherin to the Benedictine monastery in Shrewsbury and pilgrimages then centred on Holywell and Shrewsbury. The cult of St Winifred grew in prominence until, in 1415, the feast of St Winifred was ordered to become a major solemnity throughout England and Wales. The following year, *in 1416*, King Henry V made the 45-mile pilgrimage, on foot, from Shrewsbury to Holywell to give thanks for his victory at Agincourt (1415). His route would very probably have passed through the parish in which the Gowers' home lies. It should be noted that Ellis Peters used the transfer of St Winifred's bones to Shrewsbury as the inspiration for her first Brother Cadfael novel, *A Morbid Taste for Bones* (1977). David is a great fan of the Cadfael books and TV series. Rose-Mary says that she does not read the books, but did watch the series. It may also be significant that the Gowers' Labrador dog was named 'Cadfael'.

The Photographic Anomalies

I have examined prints of eleven photographs taken by Rose-Mary between March 1997 and March 2001 that may show anomalous phenomena. These are all reported to have been taken using Kodak Gold 400 or 800 film with a Konica Z-up 70 Super fully automatic compact 35 mm camera (35–70mm lens f/5.2–9.8). Rose-Mary reports that she usually uses out-of-date film that she obtains from a friend. All films are routinely taken for processing to the Tesco Store in nearby Mold. The photographs, in date order, are:–

Face in the Barn Window (Figure 1)

This was reportedly taken in March 1997, along with a large series of other photographs of their new property. Rose-Mary claims that she did not see anything unusual at the time, in which case the photographic composition might seem rather strange. The photograph appears to show a face, possibly with some kind of head-dress, almost fully filling the window at the back of the metal barn. The window normally looks out to the foliage of a tree directly behind the barn. On my visits, the window was quite heavily cobwebbed. The window glass is wire-meshed and measures approximately 50 x 75 cm. The dimensions of the 'face' are therefore about twice life-size. The eyes are notable for their asymmetry. I have requested the original negative of this photograph for scanning but unfortunately Rose-Mary has been unable to locate it.

Monk's Bench with Monk and Cross (Figure 6)

This was reportedly taken around March 2000. It apparently shows the white head and torso of a hooded figure, and a large white cross, on the back-board of the bench. The other white patches on the left and right of the bench appear to be flash reflections from the varnish. Again I have requested the original negative of this photograph, but it has not been discovered.

Monk in the Lounge Window 1 (Figure 5)

This was reportedly taken around May 2000. The photograph apparently shows a dark outline behind the lounge window of a shape that could be

Journal of the Society for Psychical Research [Vol. 66.4, No. 869

interpreted as a cloaked and hooded life-size figure. A large pottery vase that normally stands in the far left of the window (as viewed from the outside) is missing in the photograph. Unfortunately, the original negative is again unavailable.

Monk in the Mirror (Figure 7)

This was reportedly taken in September or October 2000. Rose-Mary says that she took the photograph of her digital artwork, which she had hung underneath the mirror. She was not intending to photograph the mirror, and the top of this is cut off in the photograph. However, this does beg the question: why was the photograph taken in portrait rather than landscape orientation?

The photograph seems to show the head and torso of a white, cloaked figure reflected in the mirror. Possibly the head is hooded, although the image could be interpreted as showing hair. Some people interpret the image as that of a girl rather than a monk, especially when the photograph is viewed as a negative.

Of all the photographic anomalies, this is the clearest, most dramatic, and it initially appeared possibly the most difficult to hoax. I therefore chose it for closer examination. The framed mirror itself measured 50 x 60 cm. It contained a single sheet of glass (43 x 53 cm) held with a sheet of hardboard. The back surface of the glass was silvered and printed with an Art Noveau style decoration. In the original print of this photograph (Figure 7, left) the white figure of the 'monk' seemed to disappear behind the decorative printing. This could indicate that the figure was 'in the room' and reflected in the mirror rather than being painted or daubed on the surface of the glass. In order to examine this image more closely, I borrowed the original negative, which was scanned at 2700 dpi with no manipulation, using a Nikon Coolscan III film scanner (Figure 7, right). The scan revealed that, in several places, traces of the white figure appeared *in front of* the decoration (at head, shoulder and base). This is consistent with the image being daubed on the surface of the mirror. Experiments using the mirror have shown that a similar effect may be produced using a layer of a substance such as white polish or sun-cream to create the image. Because other 'Brother Doli' phenomena involve graffiti and daubed surfaces, this is not conclusive evidence of a hoax, but it must be considered highly suspicious. I shared these concerns with David and Rose-Mary on 14th January 2001. Rose-Mary has since discovered another photograph showing her artwork beneath the mirror (with no anomalies). She now believes this is the photograph she took and is no longer certain that she took the photograph of the 'Monk in the Mirror'. The mirror itself is reported to have smashed when it fell from the wall in November 2001.

Mirror with Amorphous Shapes (not illustrated)

In November 2000, Rose-Mary took four further photographs of the mirror that were found to contain amorphous white shapes that are similar in general appearance and texture to the 'Monk in the Mirror'. On one photograph, Rose-Mary claims to be able to see a face that seems strangely familiar to her. Again there is evidence in two of the photographs that the images are daubed on the surface of the mirror.

Monk in the Lounge Window 2 (not illustrated)

Also in November 2000, Rose-Mary took another photograph of the lounge window that apparently shows a dark outline of a figure that is remarkably similar to that in the earlier photograph of the lounge window (i.e. Figure 5). Again the pottery vase is missing in the photograph. The original negative has not been located.

John-Paul in Light (Figure 11)

In February 2001, Rose-Mary took a photograph of John-Paul sitting in a chair in front of the fireplace in the lounge. She claims that as she took the photograph there was a bright flash that came from her left. The photograph shows a bright area of white light fogging the image at the left. John-Paul and the chair in which he is sitting (but *not* the rest of the room) are also bathed in a bright white light. John-Paul himself looks rather startled. From the direction of the shadows cast on John-Paul's clothing, the light would appear to be coming from the right of the photograph. This is the direction of the lounge window that faces out onto the driveway.

Figure 11. John-Paul in light.

Smoke in the Lounge (Figure 12)

Rose-Mary reports that she took this photograph in March 2001, sometime around the time of Brother Doli's announced departure (to be discussed later). It was not processed until 7th June 2001. When she showed the photograph to John-Paul, he is said to have replied that it is Brother Doli leaving. The photograph (apparently taken with flash) shows the lounge fireplace obscured by a large amorphous area of what looks like swirling white smoke or steam rising up from the direction of the floor towards the ceiling. I have examined and scanned the original negative, which itself shows no suspicious features. However, it may be significant that after two normal (holiday) photographs, three blank frames immediately precede the smoke image (No. 5a on the negative strip).

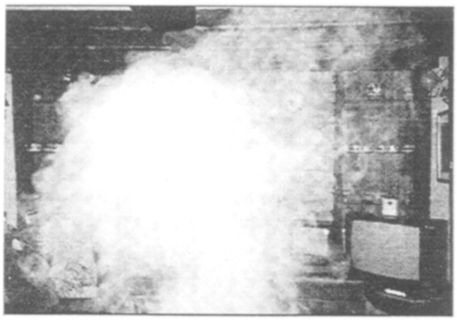

Figure 12. Smoke in the lounge.

Journal of the Society for Psychical Research [Vol. 66.4, No. 869

The Attempt to Videotape Phenomena

Because of the frequency with which new stains and carvings appeared during the course of the survey, an attempt was made to record their manifestation on videotape. For this purpose a small surveillance camera was set up to view the white-painted plastered kitchen wall adjacent to the back door (the primary route of access to the house). This location was chosen because of its convenience and unobtrusiveness, and because it would allow the monitoring of people entering and leaving the house. The wall in question was also one of the 'hot spots' where stains and (especially) carvings were regularly appearing. The resolution of the camera and recording was insufficient to show a faint stain or small carving on the wall. However, should a stain or carving appear during surveillance, the recording would hopefully provide evidence of whether or not someone had been at the wall to produce it.

The camera was an unbadged monochrome CCTV device with infra-red illumination. This used a 1/3" SONY CCD image sensor scanning at 290,000 effective pixels. The light sensitivity was 0.1 Lux. The camera gave good results in normal daylight or artificial lighting, and the infra-red capability produced an acceptable image at night with all house lights switched off. A Toshiba KV-6200E time-lapse video cassette recorder was set to record for seven days (168 hours) using a standard VHS 3-hour tape. Frames were recorded approximately every 2.25 sec. Unfortunately the set-up had no battery backup to continue recording in case of power failure, although recording would recommence when power was resumed. The VCR and camera were installed by technicians on top of a standing cabinet, just below ceiling height, with wiring neatly trailed along the walls to the nearest power socket on the other side of the kitchen. Accurate timing of recorded events was ensured by a digital on-screen display. The wall under observation also held a large clock with a second hand, providing useful confirmatory evidence of time.

Recording commenced at 14.40 on 25th January 2001 and continued until 10.35 on 1st March 2001. During this period, five tapes were recorded. I personally replaced all tapes. The intention was to do this at convenient times within the seven-day maximum period (to eliminate over-recording).

Tape 1 (14.40, 25th January – 17.06, 5th February 2001)

Tape 1 was not replaced within the seven-day recording period because Rose-Mary contracted flu. The tape automatically rewound and, as a result, is over-recorded at the beginning. No changes on the wall were noted.

Tape 2 (17.07, 5th February – 15.16, 10th February 2001)

Tape 2 was replaced on Saturday 10th February. On this date a film company visited the house to record material for a documentary on paranormal investigations in the UK. When I arrived at 11.00 am, I noticed that a small carved cross (1.5 x 2 cm) had appeared on the video wall. However, Rose-Mary reported that there had been a power cut at about 6.00 pm on Wednesday 7th February and also that the lights had flickered, but not gone out, during the same afternoon. David confirmed the power cut which occurred shortly after he had arrived home. It had lasted, they said, no more than two minutes. The

hope was therefore that a short gap in an otherwise continuous recording would confirm this. Furthermore, two minutes would seem insufficient time for anyone to have executed and repainted the new carving. However, when Tape 2 was examined, it showed a gap in the recording on 7th February from 16.20 to 18.05, i.e. a period of 1 hour 45 minutes (the clock of the VCR had a battery back-up and the timings were confirmed by the clock on the wall). Clearly this would have been ample time for the carving to be made. The simplest explanation is that the video recording had been stopped or paused at 16.20. It may also be significant that when the recording recommenced at 18.05, the position of the camera had shifted slightly, indicating some mechanical interference with the equipment (although the camera was not in direct physical contact with the VCR).

Tape 3 (15.17, 10th February – 18.21, 15th February 2001)

Although no changes occurred on the wall during the recording of Tape 3, two gaps in the recording occurred on Sunday 11th February. A six-minute break occurred between 16.58 and 17.05, followed by a longer break of 35 minutes between 17.11 and 17.46. No power cuts were reported.

Tape 4 (18.22, 15th February – 11.01, 22nd February 2001)

Tape 4 is the first to have recorded fully and continuously. The family was away for three days during this recording. No changes were observed on the wall.

Tape 5 (11.02, 22nd February – 10.35, 1st March 2001)

Tape 5 was planned to be the final attempt at recording. On Monday 26th February, Rose-Mary e-mailed to say that some brownish stains had appeared on the video wall. On Thursday 1st March, the tape was collected and the equipment uninstalled by technicians. Two smudged brown circular 'stains' were found on the wall (each approx. 10 cm diameter). These stains were, however, quite different in appearance from previous stains and looked as if they had been smeared on the surface of the wall with some brown substance. Also unlike other stains, Rose-Mary was able to remove them partially by rubbing with a cloth.

When the tape was played, a full and continuous recording was found. Close inspection of the recording reveals several occasions during the week when John-Paul makes what seem to be 'gestures' at the wall in the vicinity of the new stains. This unusual behaviour is first noted on Thursday 22nd February at 12.33. At 14.24 the same day, John-Paul is seen 'waving' at the wall for 10 sec. The following morning, at 08.08, John-Paul is seen gesturing and waving for 45 sec. Nothing unusual happens at the weekend, but on Monday 26th February at 10.02 John-Paul is seen wearing a hooded fleece gesturing at the wall with what seems to be a stick or pencil. Over a two-and-a-half minute period, John-Paul moves back and forth seven times to the wall. On Tuesday 27th February at 09.07, John-Paul again appears at the wall wearing a hooded fleece and gestures in the vicinity of one of the new stains for 45 sec. On Wednesday 28th February he is again seen briefly gesturing at the wall at 09.05, 10.04, 10.33, 10.53 and 13.30 (wearing the hooded fleece on the last four

Journal of the Society for Psychical Research [Vol. 66.4, No. 869

occasions). Finally, on Thursday 1st March, at 10.01, John-Paul is seen at the wall gesturing for 45 sec. On this occasion he again appears to have a stick or pencil in his hand. Of the 11 occasions on which John-Paul is recorded gesturing at the wall, six occur less than ten minutes after Rose-Mary is shown exiting through the back door, leaving John-Paul presumably alone in the house. All but one of the occasions when John-Paul is wearing the fleece, Rose-Mary has been recorded having left the house.

Examination of the first four videotapes shows no unusual behaviour by John-Paul at the video wall, and no occasion when he is wearing the hooded fleece. Rose-Mary reports that this fleece appeared in his room about three months earlier and that no one knows where it came from, or to whom it belongs.

My Experiences of 'Brother Doli'

On my first visit (12th November 2000) my camera case, which I had deliberately left on a kitchen unit, was found to be missing. It was discovered, after a search, on the floor behind a chair in the lounge. David, Rose-Mary and John-Paul were in the house at the time and all denied having moved it.

On 24th November 2000, at a time when there was some uncertainty about whether the investigation would proceed, I received an e-mail sent from an account (named 'mynach') used for Rose-Mary's art-related business which simply said "mynach dim dim dim" (monk no no no). According to Rose-Mary, several other recipients reported similar e-mails on or about that date.

On 14th December 2000, I received another e-mail from 'mynach' which said "tangnefedd tangnefedd mynach" (peace peace monk). It was apparently sent and received at 11.47 am, when I was at the Gowers' home surveying the phenomena. Rose-Mary and John-Paul were in the house at the time, although it is always possible that the e-mail was sent from a remote computer.

Also on 14th December 2000, I was taken into the master bedroom to record the stains there. On opening the door to the bedroom Rose-Mary and I were immediately struck by an overpowering, heavy smell, apparently of church incense. In response to this, Rose-Mary rushed to open the bedroom window to try to clear the air. As I was preparing to leave the house at about 17.15 that day, a roll of rubbings that I had made was found to be missing from the kitchen table where I had left it. A brief search was unable to locate the roll. I found the roll two visits later (7th January 2001) as I was preparing to leave the house. It had been slid over a measuring stick that I had left in the study earlier in the day. Rose-Mary and John-Paul were in the house on 14th December, while David was also present on 7th January.

As I was collecting up my things to leave about 16.50 on Saturday 10th February 2001, shortly after the film crew had left, I found a woolly hat had been placed in my camera case, which I had left on one of the kitchen units. The hat turned out to belong to John-Paul. Clearly this could have been put there by any of several people that day, including one of the four members of the film crew.

The final unusual occurrence that I can report relates to a photograph taken by me of stains on the kitchen wall that normally held the decorative mirror (removed for photographing). The photograph shows a very striking stain of an

unusual monk-like shape (approx. 22 x 36 cm) that I do not remember seeing at the time (Figure 13, top right). Unlike the other less prominent stains on the wall, it is not marked with a sticker. This particular stain was not apparent to the naked eye when the wall was later examined and neither David nor Rose-Mary can recall ever having seen it.

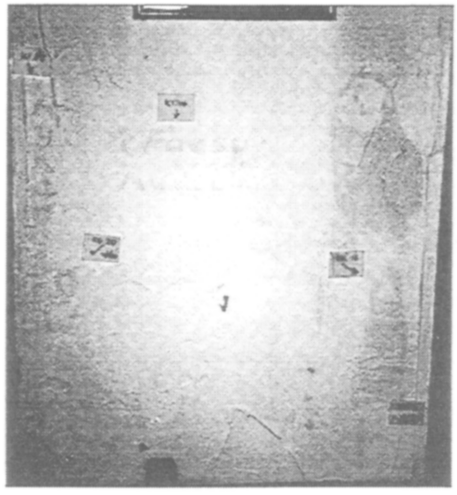

Figure 13. Unmarked 'monk' stain (top right).

Farewell Brother Doli

On Monday 26th February 2001, I received an e-mail from Rose-Mary, which reported that John-Paul had said that Brother Doli was going away to a happy, smiling place because he (J-P) was going to be 16 and was too old. On Thursday 1st March 2001 another e-mail from Rose-Mary reported that, in response to some leading questions, John-Paul had said he was helping Brother Doli and used a brown 'secret potion' to make the stains and a special black stick for the carvings. John-Paul also said that the photographs were "another film" although the meaning of this was unclear.

On Saturday 3rd March 2001, at a family gathering (which I attended) to celebrate John-Paul's forthcoming birthday, he was coaxed by his sisters into demonstrating the gesturing that he had been making at the video wall. John-Paul made a gesture similar to making the sign of the cross, although with two horizontal movements, one across the top and one at the bottom. When asked to draw it on a piece of paper, he made a shape like a capital 'I', again with horizontals at both top and bottom.

On Monday 5th March 2001, Rose-Mary e-mailed to say that the carved 'gwyn' on the video wall now reads "gwynfa" (paradise). Another e-mail later that day reported that a new stain of a cross had appeared on the same wall, and that there had been many bangs and crashes in the house. "Maybe," she mused, "it is Brother Doli's last stand!"

Journal of the Society for Psychical Research [Vol. 66.4, No. 869

Tuesday 6th March 2001 was John-Paul's sixteenth birthday. Rose-Mary e-mailed to report that John-Paul had said that Brother Doli had now gone and would not return.

On Friday 23rd March 2001 Rose-Mary e-mailed to say that about 09.20 she had seen an apparition of a heavily pregnant girl, about 12 years of age, stroking the cat on the patio. The girl was dressed in a blue cloak and a 'floaty smock'.

On 29th April 2001, Rose-Mary reported in an e-mail that many of the stains had gradually disappeared and that there had been little happening at the house other than a door continually opening and a sweet smell in and around the house.

Despite Brother Doli's heralded departure, over the next several months the family continued to report noises (bangs and crashes) in the house, as well as faint sounds that seemed to be human voices. A few stains also occasionally appeared on walls inside and outside the house. These were similar in appearance to the Welsh word stains, but simply said "Jane" and were written in a different script style. The word 'Jane' is also reported to have appeared as intrusions in some e-mails and printed documents, and as a lawn cutting. 'Jane' has also been found carved in the trunk of the holly tree and in a lounge wall.

On 13th April 2002, the word 'Tom' was discovered carved into the holly tree. On the same day, several 'Tom' stains were also found on an outside wall and on the fireplace, together with a new cross and a coronet shape. Rose-Mary reports that a few weeks earlier John-Paul had seen a boy with longish hair sitting on his bed reading his James Bond book. A week or so later, Rose-Mary was walking down the drive and reports seeing a youth with shoulder-length fair hair waving to her through the study window. Only David and John-Paul were in the house at the time.

On the basis of events reported since 6th March 2001, it would seem that both the apparitional and physical phenomena are continuing, although at a lower level of intensity than before. The form of the physical phenomena (e.g. stains and carvings) remains essentially the same, although their content no longer expresses a clearly religious agenda. Further investigation of these phenomena is clearly indicated and I intend to continue my involvement in the case. More than a year has now passed, however, since the last clear manifestation of monk-like phenomena. It may therefore be an appropriate time to conclude that the personality of 'Brother Doli' has indeed departed from the Gowers' home.

DISCUSSION

This case is both complex and multifaceted. Many of the apparently psychokinetic phenomena are consistent with a typical poltergeist-type manifestation (e.g. Bender, 1982; Gauld & Cornell, 1979; Owen, 1964; Roll, 1976, 1977). These features include object movements and disappearances, bangs and imitative noises, electrical disturbances, and pools of water. Also typical is the 'playful' and disruptive quality of the phenomena, including difficulties experienced in attempting to record their manifestation directly. Additionally, there are features in the family dynamics that are often found in poltergeist cases (e.g.

Roll, 1976, 1977; Teguis & Flynn, 1983). These include the presence of a child or adolescent (John-Paul), and preceding family disruptions (e.g. moving house and the moving away from the family home of four out of five children). A strict religious background in the family is also quite typical of poltergeist cases (Teguis & Flynn, 1983). Although this may be observed in the childhood of both David and Rose-Mary, it does not seem to apply within their own family. The fact that Rose-Mary's family apparently experienced a poltergeist when she was a child may be relevant, however.

Other aspects of the case are more unusual. Stains and especially carvings of shapes and words are rare in haunting and poltergeist cases, although there are some famous precedents. Perhaps the clearest parallel is with the controversial 'Marianne' writings on walls and paper in the Borley Rectory case (e.g. Banks, 1996; O'Neil, 2002). Similarities may also be suggested with the Spanish 'Faces of Bélmez' (e.g. Tort & Ruíz-Noguez, 1993) although the images in that case were much more painterly and naturalistic. Like the stains in the present case, however, the Bélmez faces were said not to respond to attempts at removal using detergents or by scrubbing. Also like the present stains, the faces were reported to appear, change and eventually disappear without any obvious explanation.

Another distinctive feature of the present case is the sheer volume, frequency, duration and variety of the phenomena involved. Although haunting cases may continue for decades or centuries, poltergeist manifestations typically last for only a few months, although cases spanning several years have been reported (Roll, 1976, 1977). The first clear poltergeist-type activity that may be identified in this case was the appearance of the stained cross on the fireplace in October or November 1998. At the time of writing, more than three years later, similar phenomena continue to be reported.

The present case is also notable for the absence of any bombardment by projectiles or *observed* movements of objects, which are two of the most common features of poltergeist activity (Roll, 1976, 1977). Smells, temperature changes, and visual apparitions, on the other hand, are generally more typical of hauntings rather than poltergeist manifestations (Owen, 1964; Roll, 1976, 1977). Perhaps the most unusual feature of this case, however, is the apparently benign quality of the phenomena. In contrast, poltergeist manifestations are usually annoying, unpleasant, very disruptive or traumatic, often expressing indirectly underlying emotional tensions within the family (e.g. Roll, 1976, 1977; Teguis & Flynn, 1983). Although the 'presence' of Brother Doli undeniably affected the family dynamics in various ways, his influence generally seemed to provide the family with a sense of common interest and focus. The benign nature of the manifestations is most obviously expressed in the clearly and consistently *religious* language and imagery of the stains and carvings, although the apparitions were also reported to impart a sense of peace and serenity. It is, of course, arguable whether the vision of the Virgin Mary and subsequent healing allegedly experienced by the Dooleys should be treated as connected with the other phenomena in this case. My own view, however, is that there are sufficient similarities to presume a connection.

Although some of the phenomena may be explainable in terms of natural artefacts, suggestibility of witnesses, errors of perception, or lapses in memory,

Journal of the Society for Psychical Research [Vol. 66.4, No. 869

it is obvious from the evidence of the stains, carvings and photographs that something very unusual has been going on at the Gowers' home. The question, then, is whether these unusual occurrences indicate a genuine case of paranormal activity or whether, on the contrary, they are a hoax. We should also remain open to the possibility that the case may represent some complex mixture of the genuine and fraudulent (cf. Roll, 1976, 1977).

If we consider the evidence of the physical phenomena (e.g. stains, carvings, photographs, object disappearances, e-mails) then it is clear that all of them *could* be hoaxed. Disappearing objects and e-mails would be relatively easy to hoax. Although we do not yet know the chemistry involved, the stains on plaster and stone may not be difficult to produce, although their disappearance would perhaps require a more subtle knowledge of chemistry. David, of course, has degrees in Chemistry and he readily concedes that this makes him a likely suspect in any hoax.

I remain somewhat perplexed by the prominent stain of the 'monk' shape that appeared in a photograph I had taken (Figure 13), but that did not seem to have been there at the time (nor was observed subsequently). It is, of course, possible that my memory is mistaken, or that there may have been a faint stain on the wall that was ignored during the survey and for some reason has been particularly highlighted in the photograph. On the other hand, Rose-Mary has reported similar photographic additions on several occasions. It is perhaps significant that I can provide possible independent confirmation of this.

Most of the carvings in plaster, stone and wood could easily be executed with a little care and time. The very finely executed carving of words on the 'Monk Stone' would seem, however, to be the work of a skilled craftsperson. The overpainting of the plaster carvings would also be difficult to do convincingly and undetectably, as would be the apparent disappearance and filling in of some of these carvings. On the other hand, some of the plaster carvings have appeared to show very crude attempts at filling and overpainting.

There is some evidence that strongly points to certain phenomena being a hoax. Most obvious, perhaps, are the apparent spelling mistakes that seem to be consistent with errors made by a non-Welsh-speaking person, possibly someone with poor eyesight, who has access to a Welsh dictionary. Then there is the evidence of the 'Monk in the Mirror' photograph, which seems to indicate quite clearly that the image has been daubed on the mirror.

The sceptic will also point out a number of other weaknesses or suspicious features in this case that could suggest hoaxing. These include:–

- An apparently Welsh monk, who can write Welsh words, but cannot construct a sentence in the language.
- The apparent attempt to respond to hints dropped by the researcher (e.g. for inaccessible stains, disappearing and raised carvings).
- The 'convenient' power cut during video surveillance.
- The announcement of Brother Doli's farewell that occurred within six weeks of sharing my concerns about the 'Monk in the Mirror' photograph.
- The failure, following this sharing of concerns, to locate negatives of other key photographic anomalies.
- The failure to confirm the existence of the Dooleys.

If some or all of the phenomena are a hoax, it seems clear that either or both David and Rose-Mary must be the perpetrators (possibly conspiring with others). Because of the amount of time she spends at home, Rose-Mary would appear to have the clearest opportunity. Also, many of the phenomena involve her directly (e.g. the apparitions and photographic anomalies). It is inconceivable that John-Paul has the ability to perpetrate such a sophisticated hoax unaided and the other children would seem to have little opportunity to play anything more than a minor role in any conspiracy. On the other hand, on several occasions, new stains and carvings have appeared following family trips away from home. It is conceivable, therefore, that some other person may have access to the house during these times.

Given the range of phenomena reported, a hoaxer would need to have at least a general knowledge of typical poltergeist manifestations, although this is not difficult to acquire. Rose-Mary had her own experience of poltergeist activity during her childhood. As a result of this, both Rose-Mary and David have a long-standing interest in poltergeist and related phenomena. They are also past associate members of the SPR.

The question of the possible motivation for a hoax is a relevant consideration. The family does not seem to be making substantial money from the case at this time, although it is possible that some future financial exploitation might be anticipated (e.g. a book or film). It is true that Rose-Mary's art business may benefit from the connection she has established with the Brother Doli phenomena. However, Rose-Mary's artistic work did not commence until January 2000, nearly three years after the first phenomena (the Marian visions) were reported. Another obvious possibility is that the family enjoys the interest generated in the case, especially among the media and on the Internet. Rose-Mary in particular may be said to have 'promoted' the case in various ways, through radio and TV appearances, in newspaper interviews and on several websites (e.g. Gower, 2000, and others cited in the references). I do not know the individuals or the family dynamics well enough to speculate whether there might be other, more complex, psychological motives for a hoax. Perhaps the most straightforward explanation might be that the phenomena simply represent the hoaxer's hobby, or are an expression of creative fun, with an additional reward being the delight of fooling others.

Against the argument that this is a hoax perpetrated by a single person, there is the testimony of the other members of the family confirming that strange noises are frequently heard in the house and that the stains and carvings often appear under seemingly impossible conditions. Also there are the apparition-type experiences reported by the daughters (John-Paul's experiences are, of course, very difficult to assess). This confirmation could indicate either a family conspiracy, or else that at least some of the phenomena may be genuine (i.e. not fabricated, albeit not necessarily paranormal). Alternatively, it may indicate that other members of the family have simply responded to the suggestions provided by the framing of the phenomena as the work of 'Brother Doli' and have therefore been set to experience and interpret events at the house accordingly.

If the case involves genuine poltergeist-type activity, the question arises as to who may be the *focal person* or catalyst (Bender, 1982; Owen, 1964; Roll,

Journal of the Society for Psychical Research [Vol. 66.4, No. 869

1976, 1977). John-Paul might appear to be the obvious candidate, especially considering his age together with his activities and reported statements in the final days before Brother Doli left. However, as mentioned, he seems to lack understanding of, and shows no real interest in, the phenomena and there is little evidence generally for his direct involvement. In contrast, Rose-Mary does appear to be the central figure in many of the reported events. It is therefore possible that she may be the focus and that she possesses some latent psychokinetic potential. Rose-Mary's own reported childhood poltergeist experiences may be indicative in this context.

The third main possibility is that this is a mixed case in which there is a core of genuine but relatively low-level paranormal phenomena (e.g. noises, apparitions, and possibly some staining) that has been deceptively imitated or elaborated upon. Such elaboration or 'imitative fraud' (Cox, 1961), perhaps occurring in a dissociated state of mind, has been suggested as a feature to be considered in poltergeist investigations (e.g. Bender, 1982; Roll, 1976, 1977). On this assumption, investigation becomes particularly difficult because the suspicion or discovery of some hoaxed elements does not immediately invalidate all other aspects, although it inevitably tarnishes the overall reputation of the case. The investigator's problematic task then becomes one of carefully examining each phenomenon separately in the attempt to establish or eliminate the various possibilities for fraud.

In conclusion, the phenomena that have occurred, and are continuing to occur, at the Gowers' home are extraordinary. They also remain highly ambiguous. The reader may wish, of course, to draw his or her own conclusions from the evidence I have outlined in this paper. In my opinion it is not possible, at this stage, to be certain about the status of this interesting case.

ACKNOWLEDGEMENTS

I would like to express my particular thanks to David and Rose-Mary Gower for their invitation to conduct this investigation, for their continuing openness and co-operation, and for their most generous hospitality on my visits. To their family, thanks for their willingness to discuss the case. Keith Nicholson, Mike Kavanagh, and Steve Lawler of Liverpool John Moores University contributed invaluable technical advice and assistance. Dr Branwen Jarvis of University of Wales Bangor provided most helpful consultation on the Welsh language. Three anonymous reviewers and Dr Zofia Weaver gave helpful and detailed suggestions for improvements in content and style.

School of Psychology
Liverpool John Moores University
Henry Cotton Campus
15–21 Webster Street
Liverpool L3 2ET m.i.daniels@livjm.ac.uk

REFERENCES

Banks, I. (1996) *The Enigma of Borley Rectory*. London: Foulsham.

Bender, H. (1982) Poltergeists. In Grattan-Guinness, I. *Psychical Research: A Guide to Its History, Principles & Practices*, 123–133. Wellingborough, Northamptonshire: Aquarian Press.

Brother Adolphus and his Ghostly Graffiti (n.d.) Retrieved 13 May 2002, from http://www.bbc.co.uk/so/weird/believe/doli1.shtml

Cox, W. E. (1961) Introductory comparative analysis of some poltergeist cases. *JASPR 55*, 47–72.

Gauld, A. and Cornell, A. D. (1979) *Poltergeists*. London: Routledge & Kegan Paul.

Gower, R.-M. J. (1997) Marian visions and cures in a Welsh field. *The Paranormal Review 4*, 11.

Gower, R.-M. (2000) *The Gowers and Brother Doli*. Retrieved 13 May 2002, from http://www.moldweb.co.uk/monk.htm

O'Neil, V. (2002) *Borley Rectory—'The Most Haunted House in England'*. Retrieved 13 May 2002, from http://www.borleyrectory.com/

Owen, A. R. G. (1964) *Can We Explain the Poltergeist?* New York: Garrett.

Peters, E. (1977) *A Morbid Taste for Bones*. London: MacMillan.

Roll, W. G. (1976) *The Poltergeist*. Metuchen, NJ: Scarecrow Press.

Roll, W. G. (1977) Poltergeists. In Wolman, B. B. (ed.) *Handbook of Parapsychology*, 382–413. New York: Van Nostrand Reinhold.

Teguis, A. and Flynn, C. P. (1983) Dealing with demons: psychosocial dynamics of paranormal occurrences. *Journal of Humanistic Psychology 23 (4)*, 59–75.

The Gowers' Ghost. (n.d.) Retrieved 13 May 2002, from http://www.bbc.co.uk/radio4/hometruths/brotheradolphus.shtml

The Graffiti Ghost of Treuddyn (n.d.) Retrieved 13 May 2002, from http://www.bbc.co.uk/wales/northeast/guides/weird/ghosts/ghostmonk.shtml

The Welsh Monk Story (n.d.) Retrieved 13 May 2002, from http://nationalghosthunters.com/april.html

Tort, C. J. and Ruíz-Noguez, L. (1993) Are the faces of Bélmez permanent paranormal objects? *JSPR 59*, 161–171.

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

Part II
Testing Psychic Claimants

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[8]

A BRIEF OVERVIEW OF MAGIC FOR PARAPSYCHOLOGISTS

By George P. Hansen

It has long been recognized that a knowledge of magic (e.g., legerdemain) is quite useful in investigating certain types of psychic phenomena. In the early years of psychical research, there was considerable contact between researchers and magicians (for a superb overview see Truzzi, 1983). For instance, Howard Thurston, the greatest magician of his time, witnessed table levitation by Eusapia Palladino and seemed quite convinced of its being genuine (Carrington, 1954). However, in more recent years, with the increased emphasis on laboratory work, there has been much less interaction. Nevertheless, in a 1981 poll of parapsychologists, 92 percent agreed with the statement "Magicians can play a positive role in parapsychology" and 70 percent agreed that "Magicians should be consulted by parapsychologists in setting up tests for alleged psychics" (Truzzi, 1983). Indeed a number of parapsychologists have consulted magicians in the course of their work (e.g., Beloff, 1984; Bender, Vandrey and Wendlandt, 1976; Bersani and Martelli, 1983; Crussard (Randall, 1982); Eisenbud, 1967; Haraldsson and Osis, 1977; Hasted, 1981; Roll and Pratt, 1971; Shafer and Phillips, 1982; Targ and Puthoff (Marks and Kammann, 1980)). In fact, parapsychologists Arthur Hastings, Loyd Auerbach and W. E. Cox are themselves magicians. Also, the Parapsychological Association at its 1983 meeting prepared a formal statement indicating that more cooperation with magicians should be encouraged (PA Statement on Magicians, 1984).

There are a number of reasons parapsychologists should learn about magic. A knowledge of legerdemain can give the field investigator a better chance of making a preliminary evaluation of uncontrolled, ostensible paranormal phenomena. In some cases it is possible to know whether trickery was used just given a brief description of the effect (e.g., driving a car blindfolded). In laboratory work, a knowledge of trickery can give a greater appreciation of required controls. This is especially important when developing new methods of testing gifted subjects. Even if a researcher sticks to well developed methods and tests only unselected subjects, he may be called upon to referee papers which do require a knowledge of conjuring.

I think that it is also worth noting that the major critics of parapsychology in this country (Diaconis, Randi, Hyman, Truzzi, Christopher and Gardner)have all performed magic professionally or have published in magic periodicals. In addition, Diaconis, Randi, Hyman and Truzzi have been consulted regarding federal funding of psi research (McRae, 1984). Gardner and Christopher are both very highly regarded authorities on magic. The field of parapsychology has no one comparable. My own experience indicates that very few parapsychologists are familiar with (let alone read) even the most basic works on the topic. The parapsychology students at John F. Kennedy University are a notable exception (a class, Creation of Illusions, is required of them).

A knowledge of legerdemain is also helpful when consulting magicians. Conjurors' norms and values are considerably different from those of scientists. For instance, publicly revealing the modus operandi of a trick is considered by most magicians to be a serious breach of ethics. Randi has been bitterly attacked by magicians for his exposés of Uri Geller. Thus, in some quarters, exposing such fraud is considered unethical! On the other hand, other magicians have strongly criticized mentalists for posing as true psychics (e.g., David Hoy has been posthumously attacked for this (Prince, 1983)). For in-depth discussions of some of these issues and the sociology of magic see Collins (1983), Gloye (1964) and Stebbins (1982).

A confusing, but fascinating aspect of this area is the actual belief in psi by conjurors. Gardner (1983) claims that magicians are the enemy of psychical researchers, yet Truzzi (1983) noted two polls of magicians in which over 70 percent believed in some form of psi. Many magicians seem to have misperceptions about other magicians' actual beliefs. For instance, I know two professional conjurors who claim to have been friends of David Hoy; one said that Hoy did not really believe himself to have psychic abilities, the other emphatically stated that Hoy certainly believed himself psychically gifted. Today there are a number of practitioners who believe that part of what they do is "real" (Ruthchild, 1983).

A full discussion of the value and limitations of magic and magicians in parapsychology is far beyond the scope of this paper. A number of papers could (and should) be written addressing this topic. Certainly having a magician present during an experiment is no guarantee against fraud. Indeed, there are instances in which magicians are of no use and could even be detrimental. Collins (1983) very effectively delineates some of the problems involved in working with conjurors.

Nevertheless, magicians do have a place in psi research. The first step in attaining the benefits of magicians' knowledge is for parapsychologists to become a bit more educated on the topic. In this paper I want to discuss some factors in learning about magic and give a few tips on educating oneself. Here I can provide only the briefest of introductions. To become really proficient at magic takes years, but a greater appreciation can be had with less effort. Hopefully, when more parapsychologists know more about this area, a more fruitful discussion with magicians may be possible. Thus, my goal here is to give the reader a bit of familiarity with magic so that he can start educating himself. I do want to strongly emphasize that a little learn-

George Hansen is Research Fellow at the Institute for Parapsychology, Durham, North Carolina.

ing can be a dangerous thing. It is very easy to fool oneself into thinking that one can detect trickery after reading a book or two. This is definitely not the case.

Types of Magic

There are many different specialties within magic; only some of these are of importance to parapsychologists. The various types of magic are pertinent to different areas within psychical research and the level of useful knowledge of magic will depend upon the area being investigated. For instance, a person involved in studying reincarnation or near-death experiences will have little or no need of magic or magicians. On the other hand, a person investigating a physical medium under field conditions will have a great deal of need for knowledge of conjuring. Similarly, most magicians specialize in only a few areas and the vast majority would not be worth consulting by parapsychologists.

Some of the types of conjuring include escapology, illusion, kid magic, bizarre magic, spirit effects, rope tricks, close-up, card magic, stage magic and mentalism. This list is by no means inclusive and in practice the various types blend and overlap. Of these, mentalism and close-up have the greatest implications for the psychical researcher generally, though spirit tricks and bizarre magic are relevant for certain types of investigation.

Mentalism is the branch of conjuring specializing in simulation of psychic events. Examples of these include headline prediction, cold reading, mind reading, billet reading, blindfold sight and Hellstromism or so-called muscle reading. Close-up magic, as its name implies, allows spectators to observe at very short range. In fact, many of these tricks can be performed within one foot of the spectators. These feats typically involve the appearance and vanishing of small objects such as coins or cards, increasing the size of coins and other effects virtually identical to those reported around Sai Baba by Haraldsson and Osis (1977). Some of these require skillful sleight of hand, but many take only a few minutes to learn.

Bizarre magic involves such things as decapitation, cremation and demonstration of occult powers. It is usually performed for a small intimate group rather than on stage. The following description of one effect will give the flavor of this rather obscure area: "After a talk about Charon, the legendary boatman on the river Styx who escorted the souls of the damned across to Hell, you begin. You introduce a packet of five yellowed Tarot cards, and a parchment covered with mysterious symbols. One of the cards is chosen and signed. It is placed with its mates into the vessel of earth, each card being covered by a small mound of dirt. Suddenly the lights dim and unearthly sounds echo through the room. One of the mounds of dirt catches fire—a gleaming white spark. There's an incredible flash of red light and one of the spectators shrieks and jumps to his feet clutching his arm in pain. The mage ends the ceremony quickly, turning up the lights. The spectator is examined and is seen to be branded with the mark of Charon in the form of a blood red welt. Examining the vessel, the smouldering mound of earth is found to contain the chosen Tarot card. Neat, huh?" (*Epoptica*, 1983, 206-207). Needless to say, this is not typical parlor magic. Parapsychologists or anthropologists studying psi in other cultures should at least be aware of this branch. Such information is available not only in the developed countries; in fact, African witch doctors have been known to buy materials from a British magical supply house (Booth, 1982).

There is another small specialty area that might be termed "opportunistic psychic magic." This does not involve performing a routine act, but rather taking advantage of opportunities as they arise. This brand of trickery can be especially powerful and has definite implications for parapsychology. Two works which describe this in detail are ostensibly by Uriah Fuller (1975, 1980). This can be especially insidious when the observer or experimenter thinks that he is in charge of things. Parapsychologists (e.g., Cox, 1974) have claimed that one can easily say that a phenomenon is genuine if the observer-experimenter has control of the procedure. Unfortunately it is not always easy to say what constitutes "control." Randi (1981) has clearly demonstrated that, in some cases, an experimenter thought he had been in control, but in reality it was the subject who was calling the shots.

Observing Magic

Probably the best way to learn about being deceived is to actually experience it. Watching a skilled close-up artist can be mind-blowing. I would urge all researchers to spend some time doing this. There needs to be some warning, however; the vast majority of magicians present very bad magic. Most perform the same old tricks in a very boring manner. Most lack showmanship skills and have not developed appropriate timing and misdirection. Nevertheless, there are a number of accomplished ones worth watching.

One of the most direct ways to learn magic is to become involved in the local magic scene. Most moderate size cities have groups of magicians affiliated with one of the two major national associations (International Brotherhood of Magicians (IBM) and the Society of American Magicians (SAM)). These groups usually meet monthly and can be found by contacting the local magic shop. They are largely composed of amateurs and, as such, most magic performed at the meetings is low quality. However, these groups often will have a few highly skilled performers who are well worth watching. In addition, they sometimes bring in professionals for lectures.

When a person starts to learn about magic, he will find that there are two parts to every trick. The first and most important is the "effect" (what the spectator sees). The other is the method. The student who learns only the method without first seeing the effect will almost certainly be disappointed, because most tricks are accomplished by surprisingly simple means. In actuality, very, very few tricks require sophisticated technology or exotic chemicals. The real skill comes in the presentation—subtle misdirection, control of the audience and the like. Full appreciation of this cannot be achieved by mere book learning.

Conjuring can have an especially strong impact when the tricks are presented as though they are or may be real. Reynolds (Benassi, Singer and Reynolds, 1980) posed as a psychic metal bender for several university psychology classes. His presentation was so effective that a number of the students thought that he was in league with the devil! Nor are such results limited to naive college students. Martin Johnson (1975-1977) had a magician perform during the 1976 PA convention and a number of parapsychologists suggested that he was actually using psychic means to perform his tricks!

Literature

The published material on conjuring is quite immense. In fact, there is a periodical, *Epoptica,* which is devoted to reviews of current literature and apparatus; recently it has averaged over 50 pages. In spite of this, the material is hidden to the casual observer. Even a university library with several million volumes will usually have only 10 to 20 books on magic. Magic shops have a generally restricted selection. Nearly all the best, important books are available only through mail order outlets or from dealers at conventions. Rarely are they listed in *Books in Print.* In addition to most magic publications being quite obscure, they are usually expensive. A small stapled booklet of thirty pages can easily cost $10. The prices are intentionally inflated to keep the knowledge to a restricted few. In fact, books and tricks are frequently criticized in reviews for having too *low* a price!

I would suggest that all parapsychologists obtain two or three of the large catalogs of books and apparatus. I have shown several of these to a number of people who have been quite amazed at the extent of the items available. These catalogs can serve as a healthy antidote to the notion that tricks are for kids. Abbott's and Tannen's catalogs are only $5 and each contains over 400 pages. Micky Hades publishes a number of excellent works on mentalism. During a quick glance through their catalog on mentalism, I found 20 tricks with ESP cards and J. B. Rhine's name was mentioned seven times. There are other catalogs which specialize in electronic devices for fake ESP demonstrations.

The number of books is also quite astonishing. The beginning student should be aware that there are basically two different categories of books, those for the beginners or lay public and those for the serious conjuror. Although there are thousands of different tricks, there is usually considerable overlap in what is presented in most books (this is especially true for beginners' books). The elementary works can be purchased in most well-stocked bookstores; however, these are unlikely to describe how tricks of top performers of today are actually done. The two most important books on mentalism are *Practical Mental Magic* by Theodore Annemann and *Thirteen Steps to Mentalism* by Corinda (*the* most important). Also good is *The Handbook of Mental Magic* by Marvin Kaye. These volumes do provide a good overview, but one should realize that even $1000 would put only a moderate dent in the *mentalist* material available. Attached is a very brief list of sources of information. Several others have also recently prepared suggested reading lists (e.g., Auerbach, 1983; Hoener, 1983; Webb, 1983).

In this "hidden" literature there is a considerable amount on the psychology of deception (especially notable are Bruno, 1978; Fitzkee, 1981; Maskelyne and Devant, 1946 and Randall, 1982). These works are impressive from a theoretical standpoint and can give an understanding of the topic. However, a theoretical knowledge is of little use in recognizing trickery unless one has considerable experience in direct observation and performance.

While on the topic of books, there is another source outside that of the magic fraternity, that is Dover Publications. Much to the dismay of many magicians, Dover has reprinted a number of classic books on conjuring. These are best buys and easily obtained.

Conclusion

This has been an all too brief outline of a vast area. I do hope that more researchers will take the time to learn a bit about conjuring. Most parapsychologists cannot become experts, for that takes years. But with greater knowledge, investigators would be in a better position to consult magicians. I hope this paper goes at least a little way toward that end.

A BRIEF ANNOTATED LIST OF SOURCES OF INFORMATION

CATALOGS

Tannen's (Louis Tannen, Inc., 6 W. 32nd St., 4th Floor, New York, NY 10001). An 800 page catalog of apparatus and books.

Abbott's (Abbott's Magic Mfg. Co., Colon, MI 49040). A 450 page catalog of books and apparatus.

Mickey Hades (Mickey Hades International, 110 Union St., No. 500, Box 2242, Seattle, WA 98111). Hades has several catalogs of books and supplies. An especially large selection on mentalism.

Magic Inc. (5082 N. Lincoln Ave., Chicago, IL 60625). Several catalogs available. Good selection on mentalism.

MAGAZINES

Magick (1065 La Mirada St., Laguna Beach, CA 92651). Fortnightly, usually 4 to 6 pages per issue. Devoted exclusively to mentalism.

The New Invocation (Illusions, Ltd., P.O. Box 2530, Chicago, IL 60690). Six times a year, usually 12 pages per issue. Devoted to weird and bizarre magic.

Genii (P.O. Box 36068, Los Angeles, CA 90036). Monthly, approx. 60 pages per issue. This slick magazine covers all areas of magic, contains a number of columns, announcements, etc.

The Linking Ring. This is the monthly magazine of the IBM and is available to members only. It has approx. 150 pages per issue.

M-U-M. This is the monthly publication of the SAM and is available to members only. It has approx. 50 pages per issue.

BIBLIOGRAPHY

Annemann, T. *Practical Mental Magic.* New York: Dover, 1983. (Originally published 1944).

Auerbach, L. M. "Mentalism for parapsychologists." *ASPR Newsletter,* 1983, *9* 1, 4.

Beloff, J. "Research strategies for dealing with unstable phenomena." *Parapsychology Review,* Jan.-Feb., 1984, 1-7.

Benassi, V. A., Singer, B. and Reynolds, C. B. "Occult belief: Seeing is believing." *Journal for the Scientific Study of Religion,* 1980, *19,* 337-349.

Bender, H., Vandrey, R. and Wendlandt, S. "The 'Geller effect' in Western Germany and Switzerland: A preliminary report on a social and experimental study." In J. D. Morris, W. G. Roll and R. L. Morris (Eds.) *Research in Para-*

psychology 1975, Metuchen, N.J.: Scarecrow, 1976.

Bersani, F. and Martelli, A. "Observations on selected Italian mini-Gellers." *Psychoenergetics: The Journal of Psychophysical Systems.* 1983, *5,* 99-128.

Booth, J. "Memoirs of a magician's ghost." *Linking Ring,* September, 1982, 54-57, 108.

Bruno, J. *Anatomy of Misdirection.* Baltimore: Stoney Brook Press, 1978.

Carrington, H. *The American Seances with Eusapia Palladino.* New York: Garrett Publications, 1954.

Collins, H. "Magicians in the laboratory: A new role to play." *New Scientist,* June 30, 1983, 929-931.

Corinda, T. *Thirteen Steps to Mentalism.* New York: Tannen Magic, 1968.

Cox, W. E. "Parapsychology and magicians." *Parapsychology Review,* May-June, 1974, 12-14.

Eisenbud, J. *The World of Ted Serios.* New York: William Morrow, 1967.

Epoptica. Review of *The Book of Shadows* by Randy L. Clower. 1983, No. 4, 206-207.

Fitzkee, D. *Magic by Misdirection.* Oakland, California: Magic Limited/Lloyd E. Jones, 1981.

Fuller, U. *Confessions of a Psychic.* Teaneck, NJ: Karl Fulves, 1975.

Fuller, U. *Further Confessions of a Psychic.* Teaneck, NJ: Karl Fulves, 1980.

Gardner, M. "Lessons of a landmark PK hoax." *Skeptical Inquirer,* Summer, 1983, 16-19.

Gloye, E. E. *Institutionalized Secrecy and Mass Communications; An Analysis of Popular Literature on Conjuring.* Unpublished manuscript, Whittier College, 1964.

Haraldsson, E. and Osis, K. "The appearance and disappearance of objects in the presence of Sri Sathya Sai Baba." *Journal of the American Society for Psychical Research,* 1977, *71,* 33-43.

Hasted, J. *The Metal Benders.* London: Routledge and Kegan Paul, 1981.

Hoener, G. J. "Recommended reading for the psychic entertainer." *Magick,* March 11, 1983, 1585-1586.

Johnson, M. "Some reflections after the P. A. convention." *European Journal of Parapsychology,* 1975-1977, *1,* Part 3, 1-5.

Kaye, M. *The Handbook of Mental Magic.* New York: Stein and Day, 1975.

Marks, D. and Kammann, R. *The Psychology of the Psychic.* Buffalo, New York: Prometheus, 1980.

Maskelyne, N. and Devant, D. *Our Magic.* Berkeley Heights, New Jersey: Fleming Book Co., 1946.

McRae, R. *Mind Wars.* New York: St. Martin's, 1984.

Nelms, H. *Magic and Showmanship.* New York: Dover, 1969.

PA Statement on Magicians. *Parapsychology Review,* March-April, 1984, 16.

Price, D. "Has magicdom reached maturity?" *Genii,* May, 1983, 337-338.

Randal, J. *The Psychology of Deception (Why Magic Works).* Venice, California: Top Secret Publications, 1982.

Randall, J. L. *Psychokinesis: A Study of Paranormal Forces Through the Ages.* London: Souvenir Press, 1982.

Randi, J. "More card tricks from Susie Cottrell." *Skeptical Inquirer,* Spring, 1981, 70-71.

Roll, W. G. and Pratt, J. G. "The Miami disturbances." *Journal of the American Society for Psychical Research,* 1971, *65,* 409-454.

Ruthchild, M. "Cold Reading: The reality and the illusion." *The New Invocation,* October, 1983, 195.

Shafer, M. and Phillips, P. R. "Some investigations of claims of PK effects on metal and film by Masuaki Kiyota; II. The St. Louis experiments." *Journal of the American Society for Psychical Research.* 1982, *76,* 233-236.

Stebbins, R. A. "Making magic: Production of a variety act." *Journal of Popular Culture,* 1982, *16,* 116-126.

Truzzi, M. *Reflections on Conjuring and Psychical Research.* Unpublished manuscript, 1983.

Webb, W. "Starting in mentalism—an unorthodox view." *Linking Ring,* October, 1983, 65-66.

[9]

What Is Your Counter-Explanation? A Plea to Skeptics to Think Again

JOHN BELOFF

I say it is a scandal that the dispute as to the reality of these phenomena should still be going on—that so many competent witnesses should have declared their belief in them, that so many others should be profoundly interested in having the question determined, and yet the educated world as a body should still be simply in the attitude of incredulity.

Henry Sidgwick (Presidential Address to the Society for Psychical Research, June 17, 1882)

Belief is not a matter of choice. In the end, either one is convinced by the evidence of the arguments or one is not. Long before that stage is reached, however, there is ample room for dialogue. Perhaps the evidence on which we based our belief was inadequate. Perhaps the arguments on which we relied were unsound. At all events, I think it is important that the dialogue between parapsychologists and their critics should continue. There will always be those who will say that controversy is a waste of time, that the two sides will never see eye to eye, that time spent in this way would be better spent on research. But, while I have some sympathy with this view, I regard it as short-sighted. So long as incredulity remains the typical response of the scientific community to parapsychological claims, parapsychology will be accorded a low level of priority in the competition for funding and resources, and this, in turn, will retard its progress, thereby reinforcing the initial incredulity. Is there any way out of this impasse?

There are, I suggest, two assumptions skeptics habitually make that should not go unchallenged. The first is that only the strict experimental evidence needs to be taken seriously, that any other kind of evidence can be dismissed as anecdotal and unscientific. The second assumption is that, given the antecedent improbability of the phenomena, nothing short of their becoming commonplace—which means, in effect, finding a way of producing them on demand, could ever justify our accepting them at face value. The two assumptions go hand in hand. Thus, if there were, at the present time, some unequivocally

---

This chapter is dedicated to the memory of Piet Hein Hoebens (1951-1984).

360 What is Your Counter-Explanation?

repeatable psi effect, there would be every reason to emphasize the experimental evidence because it alone would allow us to satisfy ourselves as to its validity without having to take anyone else's word for it. But, of course, if such were the case, parapsychology would no longer be the controversial science that it now is. The present controversy takes as its point of departure the fact that the disputed phenomena are too unstable and elusive to permit such an outcome. We cannot even say at the present time whether such an outcome is even theoretically possible. In these circumstances the experimental evidence takes on a very different complexion. For an unrepeatable experiment is just another unique historical event; it no longer represents a recipe for obtaining similar results as experiments ordinarily purport to do.

Hence, to rely exclusively on the experimental evidence to settle the question of the basic existence of psi is to betray a profound misunderstanding of the role of experimentation in science. Scientists do not carry out experiments with the aim, primarily, of making converts, though every successful experiment strengthens the credibility of the phenomenon under investigation. They carry out experiments in order to test hypotheses and thereby advance our understanding of the phenomena. And, if a truly repeatable experiment should prove possible, it is precisely by increasing our knowledge of the phenomena that we shall arrive at it. That is why most parapsychological researchers at the present time are avowedly "process-oriented" rather than "proof-oriented" in their work, even when their original belief in the reality of psi may have sprung from some personal experience and not from their work in the laboratory. But, however important such process-oriented research may be in the long run, I do not think we have yet reached the stage where it can be made to bear the weight of the controversy directed at the proof issue.

Meanwhile, there is a danger that exclusive preoccupation with the experimental evidence may lead us to overlook the fact that what we may be getting in the laboratory is no more than a weak, fitful, or degenerate manifestation of psi. After all, no one would ever have had recourse to the laboratory in the first instance had there not been a strong presumption that psi occurs in the outside world and has done so in every age and in every society of which we have knowledge. Obviously the laboratory affords a degree of security, rigor, and control in a way that is hardly possible in the field, but such advantages must be set against the disadvantages of dealing with psi effects whose very presence can only be detected by means of statistical analysis. Should the skeptic's attention, then, be redirected to the spontaneous evidence?

The snag is that it is then that the second assumption comes into its own. The argument is essentially that which David Hume first propounded in his famous essay on miracles.[1] A miracle, he pointed out, is, by definition, a singular exception to some law of nature. But, since the laws of nature are daily

confirmed in our experience, no mere human testimony, however imposing, could ever suffice to outweigh the reason we have for doubting it. The fatal weakness of this argument, as I see it, is that it is bound to fail when put to the test. Thus we would have no difficulty envisaging a *hypothetical* situation where we would be left in no doubt whatsoever that a paranormal event had occurred. To take a concrete, if jocular, example, we could imagine a press conference in the White House at which the President of the United States, in full view of audience, security officers, and television cameras, were suddenly to vanish and, as suddenly, reappear elsewhere. Whether any *actual* event similarly rules out the element of doubt is, of course, quite another matter, although Hume, to his credit, does try to make his argument proof against such a contingency. Thus he goes out of his way to remind the reader that, only shortly before the time he was writing, a whole series of miraculous cures had been reported from the cemetery of St. Medard in Paris. (For an account of these events, which puzzled not only Hume but even Voltaire, see Dingwall 1947, Chap. 4.) He then concedes that, other things being equal, the evidence for their authenticity must be acknowledged as overwhelming. But, precisely because they are miraculous, a rational person has no option but to reject them as spurious But, in saying this, Hume is surely mistaken. A point may be reached, as our hypothetical example showed, where no one, rational or otherwise, had any option but to believe. What, in fact, Hume has done, all unwittingly, is to furnish a *reductio ad absurdum* of his own argument. If the evidence he cites for the events in question were indeed so overwhelming, we would have to accept it, miracles or no miracles. Actually, few skeptics are content these days to base their case on the a priori impossibility of psi phenomena. We have witnessed so many upheavals in science that we are much less prone to suppose than were Hume's contemporaries that either science or common sense can tell us in advance what can or cannot be the case. Perhaps we are more ready these days to agree with St. Augustine when he pointed out that a miracle was not so much contrary to nature as contrary to what we know of nature.[2]

Let us assume, anyhow, that we are not all Humean skeptics and that we are willing to consider the possible existence of phenomena that do not meet the strongest criterion of validation, production on demand. How then are we to set about deciding rationally between belief and doubt? The main aim of this paper is to suggest an appropriate strategy. It is the following. Whenever one is confronted with a claim, from whatever source, that has certain paranormal implications, one should ask oneself what normal explanation there could be that would obviate the necessity of invoking anything of a paranormal nature. The question, then, is whether this alternative explanation is more or less plausible than the original paranormal claim. This strategy will not solve the controversy, if only because what to one person may seem a perfectly reasonable

362 What is Your Counter-Explanation?

scenario may strike another as wildly implausible and far-fetched. Nevertheless, such a strategy can, I maintain, serve to sharpen a controversy that is always liable to get out of focus.

The Problem of the Counter-Explanation

The significance of the counter-explanation was brought home to me some years ago when, in answer to a challenge, I wrote a paper for the *Zetetic Scholar* in which I discussed seven classic experiments in the parapsychological literature that I regarded as strong evidence of the existence of psi (Beloff 1980). Of the four well-known skeptics whose commentaries were also published, only one, Christopher Scott (1980), offered a specific counter-explanation for any of my seven experiments. The other critics were all content to disagree with me on general grounds. The particular experiment that Scott elected to probe was the so-called "Brugmans Experiment," which took place at the University of Groningen in Holland in 1921 (perhaps the earliest parapsychological experiment to be conducted under university auspices). Only a single special subject was involved in this experiment, and the task on each trial was to point to a square on a checkerboard of 48 squares that corresponded with the square designated as a target for that trial. Scott's suggestion was that the instructions might have allowed a certain ambiguity as to when a trial was concluded and so which square had been chosen. In that case, an overzealous observer might consistently misread the signals, and this could have accounted for the highly significant scores obtained. I am not by any means convinced that this is what *did* happen. On the contrary I would have thought it inconceivable that any experimenter (and here there were three) could be so foolish as to carry on a lengthy experiment that ran to many sessions when, all the time, there was some doubt about what response the subject was actually making—which is, after all, the one critical item of any experiment. Nevertheless, I had to concede that Scott's counter-explanation was compatible with the account we have of the experiment, and I gladly acknowledge its ingenuity and the spirit in which it was offered.[3]

Some readers may suspect that, by insisting on a counter-explanation, I am attempting to shift the burden of proof from the claimant, where it rightly belongs, onto the critic who contests the claim. I must therefore point out that, once he has published his claim, the claimant has already done everything that is initially required of him. If someone then wishes to query it, it is up to that person to state his objections. Of course no one needs to set up as a critic. There are many situations in life when the wisest course of action is simply to admit ignorance or bafflement and leave it at that. But anyone who enters the arena of controversy cannot shirk the consequences. Too often critics of parapsychology

are content to use mockery and ridicule in their efforts to bring the field into disrepute. And, since the literature of psychical research is full of fraudulent and farcical incidents, they are assured of an easy time. But in the interests of truth this temptation should be resisted. My demand for counter-explanations is a plea to forgo rhetoric and examine the facts. Even so, is it reasonable to expect a critic to be able to come forward with a counter-explanation in all cases? How can we be sure, for example, that the case we are being asked to consider has been fully and accurately reported? How do we know that there might not be somewhere a fragment of evidence that, if it were available, would cast a different light on the case as a whole? Perhaps there are telltale clues buried in the data that, if we could but interpret them, would invalidate the claim. It took, after all, some thirty years and recourse to computers before definite evidence of data manipulation was finally discovered in the famous Soal-Shackleton series. (See the chapter by Betty Markwick in this volume.) Certainly, if the demand for a counter-explanation meant putting one's finger on exactly what was amiss in a given case, we could well be accused of trying to force the issue. But it is not unreasonable to expect a critic to indicate the weak points in the evidence and state where we should look for the most promising counter-explanation.

This much, at least, is presumably now common ground between parapsychologists and their critics. Where they still cannot agree is whether a given counter-explanation is more or less credible than the alleged facts that it purports to dispose of. Understandably, most critics concentrate on the latest experiments to attract attention, where information about methods and conditions are still open to inspection. But, although the task of finding a counter-explanation in this area is becoming increasingly more difficult because, with the advent of computers and automation, so many of the pitfalls that were relevant in the days of J. B. Rhine now no longer apply, for the reasons I have discussed, I do not think this is the ground where we can yet expect a showdown. If asked what it is that makes me side with the believers rather than with the skeptics, I would have to say, I suppose, that it is less this or that particular case than the global impression I derive from my survey of the literature that something real is going on that defies conventional explanations. However, one cannot argue about global impressions, so it is necessary for me to pick out a concrete example that can serve as the basis of discussion. What should that example be? People differ with regard to what they find persuasive so I can only say that, for my part, the most convincing evidence I know in the literature of psychical research is to be found in the careers of those exceptional individuals who (*a*) appeared to possess psi ability to a very high degree and (*b*) were willing to allow themselves to be used in the interests of research. This combination has, unfortunately, become increasingly rare over the years, but there still remains a galaxy of such star subjects from whose well-documented case-studies one can choose. The one I have chosen to illustrate my theme is Eusapia Palladino.

364 What is Your Counter-Explanation?

The Case of Eusapia Palladino (1854-1918)

The case of Palladino will no doubt strike some of my readers as little short of perverse. Was she not long ago discredited, I can hear them ask, when she was caught cheating and all her so-called phenomena were shown to be mere tricks? But, as so often in this treacherous field, a closer look shows how superficial such a judgment would be. There are undoubtedly a great many blemishes in her case but, equally, there are many valuable lessons we can learn from studying it. It was true enough that she preferred to operate under cover of darkness, as was the wont among mediums of that epoch, and reserved the right to veto any condition that she did not fancy. It is true enough that she cheated repeatedly, shamelessly, and outrageously. Nor can we take too seriously her own standard excuse that she could not be held responsible for anything she might do in her trances. Too many of her tricks were too contrived and rehearsed for that. And it is true, finally, that her phenomena were notoriously ludicrous, inconsequential, and repetitious even by the standards of the familiar mediumistic repertoire of the time. But, having said all that, one then has to add that all these facts were well known to her investigators from the very outset. And yet, in spite of this, they were undaunted, and their curiosity to observe ever more remained insatiable. Moreover, these same investigators were not obscure individuals or simple-minded spiritualists; they included some of the outstanding names in European science, who had everything to lose if it could be shown that they were being duped. For the fact is that Eusapia was investigated more frequently, more intensively, and by more different qualified investigators than any other individual in the annals of psychical research. Consequently, whatever construction we care to put on this fact, we cannot evade the question of how this uneducated peasant woman from Naples who could barely write her own name succeeded for nearly 20 years in enthralling the scientific elite of Europe. For, with very few exceptions, nearly all those who studied her at first hand, however incredulous they might be to start with, came to the conclusion that her phenomena were genuine—that is to say, whatever else they might portend, they were not due to trickery—and they adhered to this conclusion to the end.

So what were these strange phenomena that she purveyed?[4] They can be divided for convenience into the following main categories: (1) Acoustic phenomena, mainly loud raps. This was a very common accompaniment of sittings with physical mediums. (2) Kinetic phenomena, e.g., the movement of small objects, the playing of musical instruments, and—a prominent feature of her mediumship—the total levitation of the séance table; or very occasionally, of the medium herself. (3) A sudden loss of weight during a séance when the medium was seated on a weighing machine. (4) Cool breezes of sufficient force to cause the window curtains to billow out into the room or the medium's skirts

to billow outwards. (5) Luminous phenomena, e.g., transient light effects, luminous apparitions, etc. And, perhaps the queerest item of the whole repertoire, (6) Materializations. These were much cruder than those associated with some of the famous materializing mediums and usually took the form of rudimentary heads or hands that would emerge from behind the curtains. Sometimes, however, it looked as if the medium herself had sprouted additional limbs, or "pseudopods," as they came to be known. Sitters would also sometimes feel themselves gripped or pinched as if by a living hand that they did not see. All the foregoing phenomena were familiar to psychical researchers of that era; indeed, Palladino's phenomena were much less spectacular than those of some other mediums. One phenomenon that does appear to be unique to her case was the jet of cool air that would sometimes be observed to emanate from a scar in her forehead, usually at the end of a séance.

Her phenomena always occurred within the vicinity of her body, although not necessarily within her reach. I know of no instance where she was able to exert an effect on a remote object or on one inside a sealed container. The arrangement at her séances, which eventually became standard, was to have at least two controllers, one on each side, to keep hold of her hands and feet, although sometimes another person would lie on the floor to hold her legs. She was seated in front of a curtained recess that diagonally cut off a corner of the room. This was known as the "cabinet" and was regarded by spiritualists as a sort of powerhouse for generating physical phenomena. Illumination varied a great deal even within a single séance, depending mainly on her whim. Sometimes the only illumination came from a bare window; but, as the sittings were always held in the late evening or at night, this was somewhat residual. Sometimes, however, artificial light was permitted, and toward the end of her career electric light was regularly used, but always shaded to give a red glow. In the more important investigations a stenographer would be employed, whose sole duty was to write down an account of every incident as it occurred, dictated by one or another of the sitters. In this way, nothing would be left to memory.

The list of those who sat with her is an imposing one. I do not think there is any doubt, however, that her most assiduous investigator, who could boast of having sat through more than a hundred of her séances, was Charles Richet of the University of Paris, a Nobel Laureate physiologist, a psychologist, and a psychical researcher. Richet has been portrayed as a rather credulous man,[5] so it is only fair to point out that, when he began his career in psychical research, he was just as skeptical as most of the other scientists of that time and even joined in the chorus of derision that greeted William Crookes, though he later had the good grace to retract his words.[6] His "credulity" seems to have consisted of an inability to doubt the evidence of his senses! But, in any case, he was by no means alone in his endorsement of Eusapia, and I must now try rapidly to cover

366 What is Your Counter-Explanation?

the principal investigations of her that were undertaken between 1891 and 1907.[7]

The first person of international standing to be converted to a belief in Eusapia's abilities was the psychiatrist Cesare Lombroso of the University of Turin, now best remembered as the founder of criminology. He and a group of scientists held sittings with her in Naples in 1891 and duly issued a positive report. The following year a second investigation, in which Richet himself participated, was carried out in Milan, under the direction of G. Schiaparelli, the head of the Milan observatory. The report that issued from this investigation included photographs of complete table levitations obtained in good light. But it is of interest, in light of what I was saying about Richet's reputation for credulity, that he was not satisfied that every possibility of deception had yet been eliminated and refused to sign the report. Richet again participated in the series of sittings held in Rome from 1893 to 1894, but it was not until after he had held sittings under his own direction and at his own residence, in the summer of 1894, that he was ready to relinquish further doubts about the phenomena. *"C'est absolument absurde mais c'est vrai,"* he used to say. These investigations took place partly at his château at Carqueiranne (near Toulon) and partly on the Ile Roubaud, a small Mediterranean island (near Hyères), where his cottage was the only dwelling and where, apart from the lighthouse-keeper and his wife, Richet, his family, and his servants were the only inhabitants, a fact that made him feel all the more confident that there could be no accomplices lurking in the vicinity.

The sittings with Eusapia on the Ile Roubaud were particularly productive of phenomena, including some of the most impressive that are attributed to her, although it must be admitted that artificial light was excluded so that the only illumination came from the windows and the night sky. The sitters, however, were all seasoned psychical researchers. Oliver Lodge and Frederic Myers of the Society for Psychical Research (SPR) were Richet's guests from England, and they were joined later by Henry and Eleanor Sidgwick. Julijan Ochorowicz, the Polish psychologist and psychical researcher, who completed the team, had already directed 40 sittings with Eusapia in Warsaw, which had impressed him. Otherwise there was just the professional stenographer to keep the minute-by-minute record of the proceedings. A very positive report on these sittings duly appeared in the *Journal of the SPR* in November 1894, written by Lodge but endorsed by Myers and by the Sidgwicks.[8] One detail is worth quoting. Lodge mentions that "raps on the table, which were frequent, were so strong as to feel dangerous: they sounded like blows delivered by a heavy mallet."

One outcome of the Ile Roubaud experience was that the SPR decided to investigate Eusapia for themselves. A series of sittings were held with her in the late summer of 1895 in Cambridge, at the house of Myers, where she stayed. They were a severe disappointment. The phenomena were sparse, and time and

again she was caught cheating. Her ruse was to free one of her hands by first bringing her two hands together and then, after a certain amount of wriggling, contrive to make the controllers on either side think that each was still holding a different hand when in fact they were grasping different parts of the same hand. Why the Cambridge sittings were such a flop is still a matter for controversy and speculation. Eusapia's supporters protest that she was uncomfortable in the stuffy atmosphere of Cambridge academic society, although Eleanor Sidgwick denies that this was so and insists that they all did everything they could to make her stay a cheerful one.[9] Others lay the blame on Richard Hodgson.[10] Hodgson had come over from Boston especially for the occasion and soon took command of the situation. Ten years previously his report on Mme. Blavatsky had been issued by the SPR. He had openly accused her of fraud on an extensive scale. Probably he was now hoping to add Eusapia to his trophies. At all events, he had heard about her mode of cheating and decided that the best way to find out more about it was deliberately to let go of a hand at certain critical moments to see what she would then do. The result deeply shocked both Myers and the Sidgwicks, who felt that they had been betrayed and even doubted whether they could any longer assume that the events they had witnessed on the Ile Roubaud were all that they seemed to be. Lodge, on the other hand, took it more philosophically. But the upshot was that henceforth it became official SPR policy not to continue working with a medium who had once been caught cheating.

The policy was curiously illogical on the part of a philosopher of Sidgwick's caliber, since it was the medium's ability not her morals that were on trial. The question, after all, was never whether Eusapia *would* cheat if the opportunity arose but rather whether she *could* cheat in the circumstances and, if the answer is that she *could* cheat, then the test was invalid in the first place.[11] Fortunately, the Continental researchers were not so easily put off, so the Cambridge fiasco did not bring her career to an abrupt end. Thus, in 1898, in Paris, a large committee under the direction of Camille Flammarion, the astronomer, went into action in Flammarion's own house. One of the incidents described by Flammarion himself involved an accordion that apparently played of its own accord while he held the end opposite the keyboard and Eusapia's hands were tightly restrained. Richet participated in these sittings but then held some further sittings for her in the library of his own house in Paris. One of those invited was the distinguished Swiss psychologist Theodore Flournoy of the University of Geneva, who has given us his own account of these sittings, which fully convinced him of the authenticity of the phenomena.[12]

The next big investigation took place in Genoa under the direction of Enrico Morselli, head of a clinic for nervous and mental diseases, who held 20 sittings with Eusapia from 1901 to 1902 and then 6 further sittings there in the

368 What is Your Counter-Explanation?

winter of 1906-1907. The latter attracted widespread publicity in the Italian press and Carrington describes them as "among the most remarkable and convincing that have ever been held," but Morselli himself thought that her powers already appeared to be waning when compared with her performance during the earlier sittings. One of the sitters at the Genoa investigation was Philippe Botazzi, head of the physiological institute of the University of Naples, who in 1907 had an opportunity to test Eusapia in his own laboratory and, as did so many others, underwent the process of having his initial incredulity converted into complete conviction.

In the same year, Lombroso came back into the picture. He organized a series of sittings in the department of psychiatry at the University of Turin that was attended by a number of medical men. It was followed by a fresh series of sittings in Turin under the direction of Pio Foà, professor of pathological anatomy and general secretary of the Academy of Sciences. His committee stated in their report: "Without the objective refutable documents which remained, we should have doubted our sense and intelligence." It is worth citing one incident that involved a strong white-wood table (2′9″ high, 3′ long, and 22″ wide, weighing 17 lbs), which, instead of just receiving the loud raps like those Lodge had described, actually broke into pieces before their astonished gaze and in what is described as "very good red light" (Carrington 1909, 106). This incident followed an announcement by the medium herself that she intended to break the table for them. (It was characteristic of her to say in advance what was going to happen.) It must, one supposes, require an extra-strong dose of skepticism to doubt evidence of that magnitude!

There were still other investigations that need not detain us, but some mention should be made of the lengthy series of sittings with Eusapia in Paris, organized by the Institut Général Psychologique between 1905 and 1908. A report on these sittings was issued under the name of Jules Courtier, secretary of the Institute. The importance of this series is twofold: Photography and instrumentation, including electrical security devices for monitoring the medium's feet, were used more extensively than in any of the previous series. Secondly, the list of sitters included such luminaries as Pierre and Marie Curie and Henri Bergson. Eusapia was also subjected to a battery of psychophysiological tests, although nothing notable emerged from this. Many of the special tests devised for her failed, and on several occasions she was caught cheating, but the report is far from being dismissive of the phenomena in general.[13]

Yet, in spite of this load of testimony and the huge number of professional man-hours it represents, in 1908 it could still be said (quite unfairly in my opinion) that scientists are the last people who can be trusted to see through a wily impostor, that only a trained illusionist can appreciate the manifold possibilities of deception and sleight of hand.[14] This objection, however, does not apply to

the investigation to which we must now turn, where all three investigators not only were themselves highly proficient in the art of conjuring but were, indeed, the acknowledged experts on the special techniques of fraudulent mediumship.

The Feilding Report

At the beginning of the year 1908, the American psychical researcher Hereward Carrington went to see Everard Feilding, then honorary secretary of the SPR. Some 22 official reports on Palladino had already appeared, but Carrington persuaded Feilding that, despite its earlier ban, it was time for the SPR to have another go before it was too late, and the consent of the Council was duly obtained. The two men took themselves off to Naples, where they were later joined by W. W. Baggally, a Council member and an expert on trickery, who had sat with Palladino but without becoming convinced that she was genuine. It was this trio who made up the "Naples Committee" and, as a testimonial to their competence for the task at hand, I cannot do better than quote the words of that distinguished scholar and erstwhile Fellow of the Committee for the Scientific Investigation of Claims of the Paranormal (CSICOP), Eric Dingwall (1950):

> I was intimately acquainted with all three investigators. Mr. Carrington was one of the keenest investigators in the United States. He had unrivalled opportunities to examine the host of frauds and fakers who flourished there, and his results led him to suppose that of the alleged physical phenomena the vast bulk was certainly produced by fraudulent means and devices, as he himself asserts in his book (*The Physical Phenomena of Spiritualism*. New York 1907, London 1908). Mr. Feilding was also a man of vast experience and one of the keenest and most acute critics that this country has ever produced. . . . Moreover his scepticism was extreme, although it was modified by an attitude of open-mindedness and an unwillingness to accept critical comments when these were not accompanied by properly adduced evidence. Mr. Baggally almost equalled Mr. Feilding in his scepticism and desire for investigation. He knew more about trick methods than his illustrious colleague and thus he was better able to concentrate on essentials. For over thirty years he had attended seances, but had come to the conclusion that rarely, if ever, had he encountered one genuine physical medium. This, then, was the committee which Eusapia consented to face.

The three men booked into the Hotel Victoria in Naples and there, in their rooms on the fifth floor, between November 21 and December 19, 1908, 11 sittings were held, the minutes of which go to make up what I am here calling "The Feilding Report" (though all three put their name to it). It was published the following year in the *Proceedings of the SPR* with a gracious foreword by

370 What is Your Counter-Explanation?

Eleanor Sidgwick (Feilding et al. 1909). It is a document that is still worth reading, not just as an account of certain extraordinary phenomena but also as a record of the moral struggle of each man with himself as he wrestled with his doubts and misgivings before finally succumbing to the only conclusion that honesty would allow. It was not until after the sixth séance, so Feilding informs us, that he himself was ready to commit himself. It is worth quoting at some length from the note he wrote on December 6, following this séance:

> My own frame of mind when starting on this investigation was that, in view of the concurrent opinion of practically all the eminent men of science who have investigated Eusapia's phenomena, it was inconceivable that they could, in turn, be deceived by the few petty tricks that have, from time to time, been detected, and that it was therefore probable that the phenomena were real. All my own experiments in physical mediumship had resulted in the discovery of the most childish frauds. Failure had followed upon failure. While, therefore, I tended to accept the general hypothesis that the facts of the so-called spiritualistic physical manifestations must, on the evidence, be regarded as probably existent, my mental habit had become so profoundly sceptical, when it came to considering any given alleged instance of them, that I had ceased to have any expectation of finding it able to bear examination. The first seance with Eusapia, accordingly, provoked chiefly a feeling of surprise; the second, of irritation—irritation at finding oneself confronted with a foolish but apparently insoluble problem. The third seance, at which a trumpery trick was detected, came as a sort of relief. At the fourth, where the control of the medium was withdrawn from ourselves, my baffled intelligence sought to evade the responsibility of meeting facts by harbouring grotesque doubts as to the competency of the eminent professors who took our places, to observe things properly; while at the fifth, where this course was no longer possible, as I was constantly controlling the medium myself, the mental gymnastics involved in seriously facing the necessity of concluding in favour of what was manifestly absurd, produced a kind of intellectual fatigue.
>
> After the sixth, for the first time, I find that my mind, from which the stream of events has hitherto run off like rain from a macintosh, is at last beginning to be capable of absorbing them. For the first time I have the absolute conviction that our observation is not mistaken. I realize, as an appreciable fact in life, that, from an empty cabinet I have seen hands and heads come forth, that from behind the curtain of that empty cabinet I have been seized by living fingers, the existence and position of the very nails of which could be felt. I have seen this extraordinary woman sitting visible outside the curtain, held hand and foot by my colleagues, immobile except for the occasional straining of a limb, while some entity within the curtain has over and over again pressed my hand in a position clearly beyond her reach. I refuse to entertain the possibility of a doubt that we were the victims of a hallucination. I appreciate exactly the fact that ninety-nine people out of a hundred will refuse to entertain the possibility of a doubt that it was anything else. And, remembering my own belief of a very

> short time ago, I shall not be able to complain, though I shall unquestionably be annoyed, when I find that to be the case [Feilding 1909, 461-462].

I sometimes think that the Feilding Report should be made mandatory reading for any would-be member of CSICOP. Only those whose skepticism survived this ordeal intact would then need to apply! Yet, as this passage shows, Feilding knew only too well what to expect from his critics, and it could have come as no surprise when Frank Podmore, always the implacable enemy of the physical phenomena, wrote to the *Journal of the SPR* to attack the report. In his letter, published in the December 1909 issue, he reverts to the possibility of Eusapia surreptitiously freeing a hand or foot and thereby accomplishing all the phenomena therein described (Podmore 1909). He was answered at length by Baggally (1910) in the following issue, February 1910, but alongside further letters supporting Podmore's criticisms. In his book *The Newer Spiritualism* Podmore (1910, Chap. 4) enlarges on this theme. He minutely dissects various séances in an effort to show where Eusapia might have been able to seize the advantage, but it is something of a tour de force and made no impact on the authors of the report. Podmore, we must remember, had never sat with Eusapia. In the end, he graciously admits that the job could not have been done more competently than it was by the Naples Committee, but he took his stand on his contention that no one, given the constraints within which they had to operate, could have got the better of Eusapia; and he urged that, unless she could produce effects inside sealed containers, the SPR would be well advised to leave her strictly alone.

Although many skeptics will sympathize with Podmore's attitude, Eusapia's reputation would have been safe after the publication of the Feilding Report had it not been for the subsequent debacle of her American tour. For, in the following year, 1909, Hereward Carrington, emboldened by the success of the Naples Committee, arranged for Eusapia to visit the United States. It was a decision he would live to regret. She arrived amid a blaze of publicity and was at once in demand from all quarters. According to Carrington (1913; 1954) some 31 sittings were held in all, including the 4 that took place at Columbia University in New York. She was, he alleges, tired, overworked, and ill at ease among investigators with no previous experience of dealing with mediums. Although her sittings were by no means unproductive, they were disappointing by previous standards and, as so often happened with her when things were not going well, she lapsed into her old bad habit of cheating in the usual fashion. This was all that was needed for the skeptics to raise the hue and cry. Eminent psychologists, like Hugo Münsterberg of Harvard and Joseph Jastrow of the University of Wisconsin, after just a few sittings, rushed into print in the popular magazines to denounce her and claim credit for exposing her, oblivious to the 20 years of

372 What is Your Counter-Explanation?

patient work that had been done with her in Europe. Mark Hansel quotes at length from these outbursts as demonstrating the superior wisdom and discernment of the unbeliever and, today, if Palladino is still mentioned by psychologists, it was Münsterberg and Jastrow who said the last word on her, not Feilding, Baggally, or Carrington.

Conclusions

Are we then justified in dismissing the career of "this extraordinary woman," as Feilding called her, with the one simple word "trickery"? Trickery is, of course, another of those convenient open-ended and slippery concepts that, no less than the concept of the paranormal itself, can be invoked to explain anything whatsoever. All the same, it is not unreasonable to ask what conjurer would agree to perform under the conditions to which Eusapia regularly submitted? These included (1) performing in a private room that was first searched and then locked and sealed against intruders who might act as accomplices; (2) undergoing a thorough body search before the sitting began;[15] (3) allowing one's arms and legs to be held by sitters whose one duty was never to let go; (4) producing sometimes phenomena in light sufficient, it was said, to read the small print in one's Baedeker. At all events, one leading American magician who observed her during her American tour was so impressed by her performances that he offered to donate $1,000 to charity if any of his fellow magicians could do as much under comparable conditions. There were no takers. (See Rogo 1975, 27.)

Apart from sleight-of-hand trickery, the only other counter-explanation that was occasionally invoked in connection with Palladino was hallucination, especially with reference to her materializations and pseudopods.[16] It is a somewhat desperate explanation, especially as there was always more than one sitter; but, in the nature of the case, it is very difficult to deny. However, skeptics can always rely on one supreme ally: the poverty of the human imagination. Some things go so far beyond our familiar experience and are so inherently hard to credit that even to contemplate them imposes a severe strain on our intellectual equanimity. It is not, therefore, surprising that people are content to clutch at any straw as an excuse for not having to take these things seriously and are pathetically grateful to critics like Hansel whom they can cite in self-justification. To such people Eusapia's mismanaged American tour came as a godsend.

But, for those who are willing to make that leap of the imagination that is required if we are to put ourselves in the shoes of those who came face to face with her phenomena, who, so to speak, had their noses rubbed in them, there are a number of useful lessons we can learn from her career. In the first place, it reminds us that fraud can go hand in hand with genuine psychic ability, so that

it is always risky to generalize from the discovery that cheating has occurred. There may be all kinds of psychological reasons why certain persons in certain situations indulge in trickery. We can also learn from her case that the more fantastic phenomena are not necessarily any less real than those of lesser magnitude. In particular, we can no longer justify dismissing materialization as too preposterous to warrant serious consideration.

This last point, in turn, reopens a host of other controversial historical cases that now demand to be looked at with a fresh eye and the received opinion if necessary challenged. What about the young Florence Cook, for example? Must we ignore her because in middle age she took to cheating? Must we continue to defame the memory of one of Britain's most illustrious (not to say bravest) scientists by insisting that he was either the dupe of this 16-year-old girl or else her lover, who for the sake of sexual favors betrayed his fellow scientists and the cause of truth?[17] And what of those cases that flourished in the decade after Palladino? What about poor "Eva C," with her ectoplasmic faces that looked to all the world so suspiciously two-dimensional? The SPR investigators could prove nothing against her, though they were reluctant to authenticate her phenomena; but Gustave Geley, of the Institut Métapsychique in Paris, insists that he watched the faces taking shape and was able to produce a certain amount of photographic evidence to support his contention.[18] And must we still ignore the mountain of evidence that exists on the Margery mediumship because of one compromising incident late in her career? After all, did not the great Houdini stake his reputation on exposing her and fail ignominiously?[19] To take one more example, the most fantastic of them all, what are we to make of the Brazilian medium Carlos Mirabelli? If the so-called Santos Report were to be credited, he would rank as by far the most powerful medium on record, surpassing D. D. Home.[20] Deceased individuals known to the sitters are said to have materialized in broad daylight while the medium was seated in full view strapped to a chair. These same entities conversed with the sitters, submitted to a medical examination on the basis of which they were pronounced anatomically perfect, allowed themselves to be photographed and later dissolved into nothingness while the sitters looked on incredulously. The witnesses, moreover, were mainly persons of professional standing, many of them physicians. The only objection one can raise in Mirabelli's case is that he was never tested outside Brazil. Plans to bring him to Europe fell through. It is, however, embarrassing to have to base one's counter-explanation on the assumption that what happens in Brazil need not be taken seriously and that Brazilians, even medically qualified ones, cannot be trusted to tell the truth. The point is, however, that, even in such an extreme case as this, where we have every reason to query the evidence and to express suspicion, we cannot escape the responsibility of putting forward *some* counter-explanation.

374 What is Your Counter-Explanation?

When cornered, there is always one last trump-card that the skeptic can brandish in the face of the believer: *Where are all these marvels now?* The harsh fact of the matter is that there are no more Palladini. For all I know to the contrary, the phenomenon of materialization may be extinct and may never recur. Except in connection with poltergeist cases, we cannot even be sure that there are any more strong phenomena. Hence, since historical cases can never compete in credibility with cases that are still open to further investigation, the skeptic cannot be faulted if, like Podmore, he prefers to suspend judgment pending the advent of more compelling evidence, always provided he does not invoke spurious reasons for rejecting what evidence there is. But, equally, if the negative option is still valid, so is the positive option. It is no less rational, and it is certainly more adventurous, to adopt an attitude of basic belief as one's working hypothesis. If we remain entrenched in a rigidly conservative stance, we are apt to neglect phenomena that may be both real and important. Moreover, the basic believer is spared those intellectual contortions to which a skeptic is now driven when confronted with evidence for which there is no plausible counter-explanation. Lastly, if in the future new cases of a spectacular nature should arise, the basic believer will be in a better position and better prepared to deal with them.

Acknowledgment

I am specially grateful to the late Piet Hein Hoebens for criticizing an earlier draft of this paper, as a result of which the paper has been completely rewritten.

Notes

1. David Hume, *An Enquiry Concerning Human Understanding*, 1748, Sect. 10, Of Miracles.

2. St. Augustine, *The City of God*, 21:8

3. Prompted by Scott, Piet Hein Hoebens of Amsterdam was able to obtain from Groningen an earlier version of the report, where, sure enough, this ambiguity was more apparent.

4. The most comprehensive account known to me is Carrington (1909), but see also entry under "Paladino" in Fodor (1966 [1934]) and Dingwall (1950, Chap. 5). Dingwall's essay is still the best general introduction that I know and has an invaluable bibliography.

5. Most recently by Ruth Brandon, who in her caustic book (Brandon 1983, 135) speaks of "his will to believe and his disinclination to accept any unpalatable contrary indications."

6. See Richet (1922, 35). His long treatise is dedicated to William Crookes and Frederic Myers. An account of Richet's involvement in psychical research is given in Fodor (1966 [1934]) under "Richet."

7. Much of this is taken from Carrington (1909, Part 3).

8. Lodge (1894) writes: "Any person without invincible prejudice who had the same experience would have come to the same conclusion, *viz*: that things hitherto held impossible do occur." Lodge never saw any reason to retract this conclusion and, in his autobiography (Lodge 1931), he amplifies his account of his experiences on the Ile Roubaud.

9. In a footnote to her review of Morselli's *Psicologia e Spiritismo* in the *Proceedings of the SPR*, 21 (1909): 522. She was even then not yet satisfied as to the authenticity of the phenomena.

10. This is certainly the view taken by Cassirer (1983a).

11. It is curious how often this point has been misunderstood. Thus Ruth Brandon (1983) writes: "In all other scientific fields to be caught out just once in fraud is to be instantly discredited." But this confuses the experimenter and the subject. An experimenter who cheats is instantly discredited in parapsychology as in all other sciences and all his results discounted as suspect. But, if a subject cheats, this shows only that the experimenter has been careless and should try harder.

12. See Flournoy (1911, Chap. 7). He mentions there that "Myers was this time—as were all the others—absolutely convinced of the reality of the phenomena" (p. 246).

13. A lengthy review of this report by Count Perovsky-Petrovo-Solovovo appeared as a supplement to the issue of SPR *Proceedings* that contains the Feilding Report (Feilding et al. 1909). Flournoy is rather more brusque with the Courtier Report, which he accused of prevarication. Since they appeared incapable of saying either *oui* or *non* when it came to the authenticity of the phenomena, Flournoy (1911, 272) suggested that they ought to reply in chorus: *Nouin!*

14. This view was first, I believe, put about by the celebrated illusionist, J. N. Maskelyne. He and his son had been invited to participate in the Cambridge sittings of 1895 and he soon formed a very poor opinion of Eusapia and perhaps an even poorer one of her investigators. "No class of men can be so readily deceived by trickery as scientists" he asserted. "Try as they may they cannot bring their minds down to the level of the subject and are as much at fault as if it were immeasurably above them" (cited in Brandon 1983, 138). Feilding, on the other hand insisted that scientists, as a class, are far more reluctant than conjurers to acknowledge any sort of supernormal force (see Feilding et al. 1909, Final Note). In fact, Maskelyne, himself, once admitted to a journalist who interviewed him that, as a result of personal experiences with friends, he had to admit that there was something we could not explain about "table-turning" phenomena, though he felt sure it was not the action of spirits (Brandon 1983, 166).

15. Hansel (1980, 61) suggests that Eusapia might have taken advantage of her female prerogative to refuse such a search. However, the task was usually given to female sitters, such as the two American ladies who were invited to attend the eighth séance of the Naples series, but Carrington mentions specifically that at the conclusion of the very successful sixth séance, Eusapia made no objections to letting the three men search her (Feilding 1950).

16. Alice Johnson, at one time honorary secretary of the SPR, would occasionally advance this hypothesis when all else failed; but for a recent discussion of the "pseudopods" phenomenon, see Cassirer (1983b).

376 What is Your Counter-Explanation?

17. This was the thesis of Trevor Hall's (1962) book and is the one favored by Ruth Brandon (1983, Chap. 4). Those who are not persuaded that Sir William Crookes, O.M., was either an imbecile or an unprincipled blackguard may wish to consult the analysis of the "Katie King" séances provided by Dr. G. Zorab (1980), a Dutch scholar, although his book has so far appeared only in Italian.

18. A re-evaluation of the case of Eva C. (Marthe Béraud) is given by Brian Inglis (1984, Chap. 4).

19. Margery (Mrs. Mina Crandon) was one of the most controversial mediums of the twentieth century. Inglis (1984, 167) reproduces a photograph of her encased in Houdini's fraud-proof box ("like an old-fashioned steam-bath with a hole for her neck and two for her arms"). Houdini is shown holding one of her hands. The séance had no sooner started when the lid burst open and the phenomena continued! Mrs. Marian Nester, of the American SPR, a daughter of Dr. Mark Richardson, Margery's chief investigator, is preparing a new book about this medium that should provide new grounds for reconsidering her case.

20. See Dingwall (1930). I am indebted to Brian Inglis for drawing attention to this document in his discussion of Mirabelli's mediumship. See Inglis (1984, 221-227). Further information on Mirabelli is provided in Playfair (1975, Chap. 3).

References

Baggally, W. W. 1910. Discussion of the Naples Report on Eusapia Palladino. *Journal of the SPR,* 14:213-228.

Beloff, John. 1980. Seven evidential experiments. *Zetetic Scholar,* 6:91-94, 116-120.

Brandon, Ruth. 1983. *The Spiritualists: The Passion for the Occult in the 19th and 20th Centuries.* Buffalo, N.Y.: Prometheus Books.

Carrington, Hereward. 1909. *Eusapia Palladino and her Phenomena.* London: Werner Laurie.

———. 1913. *Personal Experiences in Spiritualism.* London: Werner Laurie.

———. 1954. *The American Séances with Eusapia Palladino.* New York: Garrett.

Cassirer, Manfred. 1983a. Palladino at Cambridge. *Journal of the SPR,* 52:52-58.

———. 1983b. The fluid hands of Eusapia Palladino. *Journal of the SPR,* 52:105-112.

Dingwall, E. J. 1930. An amazing case: The mediumship of Carlos Mirabelli. *Journal of the American SPR,* 24:296-306.

———. 1947. *Some Human Oddities.* London: Home and Van Thal.

———. 1950. *Very Peculiar People.* London: Rider.

Feilding, E., W. W. Baggally, and H. Carrington. 1909. Report on a series of sittings with Eusapia Palladino, *Proceedings of the SPR* 23:306-569 (reprinted in E. Feilding, *Sittings with Eusapia Palladino and Other Studies* [with an introduction by E. J. Dingwall], University Books, New Hyde Park, N.Y., 1963.)

Flournoy, Theodore. 1911. *Mélanges de Métapsychique et de psychologie.* Geneva and Paris.

Fodor, Nandor. 1966 [1934]. *An Encyclopeadia of Psychic Science.* New Hyde Park, N.Y.: University Books.

Hall, Trevor H. 1963. *The Spiritualists: The Story of Florence Cook and William Crookes.* New York: Garrett 1963 (reprinted as *The Medium and the Scientist: The Story of Florence Cook and William Crookes,* Prometheus, Buffalo, N.Y., 1985).

Hansel, C. E. M. 1980. *ESP and Parapsychology.* Buffalo, N.Y.: Prometheus Books.

Inglis, Brian. 1984. *Science and Parascience: A History of the Paranormal, 1914-1939.* London: Hodder and Stoughton.

Lodge, Oliver. 1894. Experience of unusual physical phenomena occurring in the presence of an entranced person (Eusapia Palladino). *Journal of the SPR,* 6:306-360.

———. 1931. *Past Years.* London: Hodder and Stoughton.

Playfair, Guy. 1975. *The Flying Cow.* London: Souvenir Press.

Podmore, Frank 1909. The report on Eusapia Palladino. *Journal of the SPR,* 14:172-176.

———. 1910. *The Newer Spiritualism.* London: Fisher and Unwin.

Richet, Charles. 1922. *Traité de Métapsychique.* Paris: Alcan (English translation: *Thirty Years of Psychical Research,* London, 1923).

Rogo, D. Scott. 1975. Eusapia Palladino and the structure of scientific controversy. *Parapsychology Review,* 62:23-27.

Scott, Christopher. 1980. Comment on Beloff's "Seven Evidential Experiments." *Zetetic Scholar,* 6:110-112.

Zorab, George. 1980. *Katie King Donna o Fantasma?* Milano: Armenia Editore.

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[10]

The Appearance and Disappearance of Objects in the Presence of Sri Sathya Sai Baba

ERLENDUR HARALDSSON AND KARLIS OSIS[1,2]

ABSTRACT: During three field trips to India to study claims suggestive of psi phenomena the investigators were able to observe at close range some unexplained occurrences which took place in the presence of Sri Sathya Sai Baba. Although no conclusions can be reached on the phenomena observed and described in this account because they occurred under informal conditions, it seemed worth while to report the events because of the challenge they offer to carry out further studies of this well-known Indian religious leader under well-controlled experimental conditions.

INTRODUCTION

Ostensibly paranormal appearances and disappearances of objects have been reported in various cultures. The phenomenon consists of an object appearing or disappearing in circumstances where no physical cause of the event can be detected. In cases where paranormal creation of the object is assumed, the process is usually referred to as "materialization." When an already existing object is "brought" by paranormal means from one place to another without visible means of travel, the phenomenon is called "teleportation" and the object is referred to as an "apport." Teleportation is said to occur in poltergeist cases (Bender, 1969; Owen, 1964; Roll, 1974). Materializations of human forms have been reported in the presence of mediums (Carrington, 1954; Hannesson, 1924; Richet, 1923; Schrenck-Notzing, 1920). Indian popular literature describes appearances of inanimate objects, usually amulets made of precious materials and said to have magical properties such as providing protective contact with a guru (Yogananda, 1969).

The appearance and disappearance of objects is of course one of the favorite illusions created by stage magicians. With the help of

1 We wish to express our gratitude to Sri Sathya Sai Baba for his kind cooperation in this investigation.

2 This research was financed through the A.S.P.R.'s James Kidd inheritance fund and by an anonymous donor in Iceland to whom we are grateful.

astonishing dexterity, diversion of attention, and some gadgetry, objects have "appeared" and "disappeared" on the magic show stage without any detection of the tricks of the trade by the audience. Enterprising showmen throughout recorded history have produced "spirits" and "demons" in religious settings in their claim to demonstrate "supernatural" phenomena.

On close scrutiny the bulk of the claims for materialization and teleportation have been explained in quite natural (and sometimes entertaining) ways (Carrington, 1920). Nevertheless, there are a few reports which keep the question open, e.g., Crookes' (1874) report on Florence Cook and D. D. Home, and the reports on Rudi Schneider by Lord Hope (1933) and Schrenck-Notzing (1920). More recently, Eisenbud's (1967) observations of Ted Serios suggest some kind of materialization interfering with light in photographic and television processes.

In spite of considerable research done in this area by psychical researchers early in this century, claims of materialization and allied phenomena have generally been frowned upon and rejected by nearly all present-day parapsychologists (Eisenbud, 1975). We too shared this point of view and did not give serious consideration to such phenomena until our encounters with Sri Sathya Sai Baba.

Sathya Sai Baba, age 51, is a religious leader who has a large following and lives in the State of Andhra Pradesh in southern India. He is not only credited by his followers and many others with a variety of psychokinetic powers (such as materializations, teleportations, and healing) but also with various forms of extrasensory perception and out-of-body projections collectively perceived. Several popular books have already been published about him, all of which deal in part with his paranormal phenomena (Kasturi, 1973-4; Murphet, 1971; Sandweiss, 1975; Schulman, 1971). We first encountered reports of these phenomena when researching deathbed visions in India in 1972-73.

During two subsequent visits to India in late 1973 and early 1975 we met with Sai Baba several times and also had lengthy interviews with a number of persons, including Indian research scientists, who had observed or experienced psychic phenomena of various kinds which they attributed to him. E.H. made another visit to India in January, 1976, for further observations and interviews with Sai Baba. In a series of interviews, we had the opportunity to discuss these matters with Sai Baba himself and to observe him in action. He speaks English but often prefers to use an interpreter. He tends to belittle the significance of his own psychic phenomena, calling them "small items," and he repeatedly stresses the importance of spiritual and ethical issues. Our bid for formal experiments was rejected with the comment that he would only use his paranormal powers for

religious purposes such as helping his devotees when they are in dire need or for invoking faith in hitherto agnostic persons, but never for purely demonstrative purposes. During the 11 interviews we had with Sai Baba, however, he did spontaneously display a number of the same phenomena for which he has become famous in India.

Interviews with Sai Baba are generally not prearranged. People who want to meet him—usually several hundred—gather outside his residence in the ashram. Twice a day he makes his rounds for a short while and chooses those he wants to see. Many wait for weeks in vain and are never granted an interview either in a group or in private. Sai Baba's interview room, where most of the phenomena we are reporting occurred, is bare, with concrete walls and floor and without carpets or any decorations. The only furniture in the room was one armchair. During our interviews we all sat crosslegged on the floor. The number of persons present with Sai Baba varied from only E.H. and K.O. to about nine persons. We observed some 21 appearances and disappearances of objects at close range, but none under controlled conditions. We shall describe four instances of these phenomena and then attempt a very tentative evaluation of their genuineness.

THE INCIDENTS

1. *Appearance of a "Rudraksha"*

The first of these phenomena concerns the possibly paranormal appearance of a "rudraksha," which is similar to an acorn, about an inch in diameter, and with a fine texture like an apricot stone. First Sai Baba presented us both with some "vibuti" (holy ash, which is probably comparable symbolically to bread and wine in Christianity). He gave us the vibuti after a typical wave of his right hand, palm down, in small circular movements that lasted two or three seconds. After a short discussion he presented one of us (K.O.) with a large gold ring,[3] again after having waved his hand in a typical manner.

While we were arguing with Sai Baba about the value of science and controlled experimentation, he turned the discussion to his favorite topic, the spiritual life, which in his view should be as "grown together" with ordinary daily life as a "double rudraksha." We did not understand this term nor could the interpreter translate it. Sai Baba seemed to make several efforts to make its meaning clear to us until he gave up and with some signs of impatience closed his fist and waved his hand. He then opened his palm and showed us a

[3] A goldsmith later examined this ring and found that it is made of gold. It was appraised at $100.

double rudraksha, which we are told by Indian botanists is a rare specimen in nature like a twin orange or twin apple.

We observed Sai Baba closely all the time we sat on the floor. After we had admired the rudraksha, Sai Baba took it back in his hand and, turning to E.H., said he wanted to give him a present. He enclosed the rudraksha between both his hands, blew on it, and opened his hands toward E.H. In his palm we again saw a double rudraksha, but it now had a golden ornamental shield on each side of it. These shields were about an inch in diameter and held together by golden chains on both sides. On top of the shield was a golden cross with a small ruby affixed to it. Behind the cross was an opening so that this ornament [4] could be hung on a chain and worn around the neck. Many, but not all, of the ornaments which Sai Baba presents to people are said to be made of precious metals and stones.

Sai Baba wears a one-piece robe with sleeves that reach his wrists. We watched his hands very closely and could not see him take anything from his sleeves or reach toward his bushy hair, clothing, or any other hiding place.

It was not possible for us to examine Sai Baba's clothing. On one occasion, however, we had an opportunity to examine two robes he had worn. He reportedly always wears robes of the same sort and when they start to wear out he gives them away. The two we examined contained no pockets of any kind or any signs of magician's paraphernalia having been attached.

We became acquainted with a former professor of chemistry in Bangalor, Dr. D. K. Banerji. One day Sai Baba visited him and his wife unexpectedly and "produced" some objects for them, as he does almost everywhere he goes. As he retired for the night in their house, he asked Mrs. Banerji to wash his robe, which she did. On this occasion she, Dr. Banerji, and a colleague of his, Dr. P. K. Bhattacharya (doctorate in chemistry from Illinois), carefully examined Sai Baba's robe and found that it had no pockets. Dr. Banerji was formerly the director of the Department of Organic Chemistry at the All India Institute of Science, which is a leading research institute in India. These three persons reported this incident to us during two independent interviews.

In an interview during his third visit to Sai Baba, E.H. repeatedly saw the sun shine through Sai Baba's thin, silken sleeves as he was sitting on a chair approximately five to six feet away from him. The

4 A goldsmith later examined this ornament and found that it contains 22-carat gold. Its value was appraised at $80. The small ruby was examined by the Gem Testing Laboratory of the London Chamber of Commerce and Industry. Because of the closed setting behind the stone it was not possible to determine whether it is a natural or a synthetic ruby. A botanist's microscopic examination of the rudraksha showed it to be a genuine example of its species.

late afternoon sun was shining through the window of the interview room where a few people were sitting on the floor around him. Most of the interview was spent on a discussion, but Sai Baba also produced a few objects which he gave to those present. As Sai Baba was sitting on a chair his arms were approximately at the head-level of those sitting on the floor close to him. The sun shining through Sai Baba's sleeves did not reveal any shadows that might indicate the presence of hidden objects. Sitting that close to Sai Baba, E.H. could several times see up his sleeve, which appeared to be empty.

2. *Disappearance of a Picture from K.O.'s Ring*

This episode concerns the gold ring that Sai Baba had presented to K.O. during our first visit. This ring had a large enameled picture in color of Sai Baba encased in it. The picture was of oval shape, about 2 cm long and 1½ cm wide, and was framed by the ring. The edges of the ring above and below the enameled picture, together with four little notches that protruded over it from the circular golden frame, kept it fixed in the ring. Thus the picture was set as firmly in the ring as if it and the ring were one solid article.

In an interview during our second visit when we tried to persuade Sai Baba to participate in some controlled experiments, he seemed to become impatient and said to K.O., "Look at your ring." The picture had disappeared from it. We looked for it on the floor, but no trace of it could be found. The frame and the notches that should have held the picture were undamaged; we examined them afterwards with a magnifying glass. For the picture to have fallen out of the frame, it would have been necessary to bend at least one of the notches and probably also to bend the frame at some point, but neither had been done. Another alternative would have been to break the picture in the ring so that it would fall out in pieces.

When Sai Baba made us aware of the picture's absence we were sitting on the floor about five or six feet away from him. We had not shaken hands when we entered the room and he did not reach out to us or touch us. As we sat cross-legged on the floor, K.O. had his hands on his thighs and E.H. had noticed the picture in the ring during the interview and before this incident occurred. E.H.'s first reaction was that the picture had suddenly become transparent. Two persons, Dr. D. Sabnani from Hong Kong and Mrs. L. Hirdaramani from Ceylon, whom we had met for the first time during the interview, certified that they had observed the large golden ring with Sai Baba's picture on K.O.'s left hand before the picture disappeared. When the picture could not be found, Sai Baba somewhat teasingly remarked, "This was my experiment."

During our next interview, which took place two days later, Sai Baba asked K.O. if he wanted the picture back, to which K.O. replied

that he did. On Sai Baba's demand, K.O. gave him the ring which he took in his hand and asked, "Do you want the same picture or a different one?" "The same," K.O. replied. Sai Baba then closed his fingers around the ring in his palm, brought it to about six inches from his mouth, blew at it lightly, and then stretching his hand toward us, opened it. In it was a ring. The enameled picture was like the one that had been framed in the first ring; the ring itself, however, was different. The first incident, the disappearance of the picture, was obviously more evidential than was its reappearance, about which there is not much we can say.

3. *Ring and Necklace for Mr. and Mrs. Krystal*

During the aforementioned interview we observed an interesting phenomenon. A lawyer from Los Angeles and his wife, Mr. and Mrs. Krystal, were present with us. Their 33rd wedding anniversary was around that day and Sai Baba seemed to be happy about the occasion. He waved his hand and as he opened his fist we saw a golden ring. He handed it to Mrs. Krystal, telling her to put it on one of her husband's fingers, as is customary for the bride to do at a traditional Indian wedding. Sai Baba's open hand was still stretched out in the air without having touched his clothing or any object. We watched closely. Immediately thereafter Sai Baba waved his hand again for two or three seconds, turned palm down, and quickly closed it. His arm was approximately horizontal to the ground, which was not a position favorable for slipping something out of his sleeve by means of gravity. We observed at close range as Sai Baba loosened the grip of his fist so that he could hold a large, bulky necklace in his hand. Its double length was about 20 to 29 inches and it contained a variety of different kinds of stones interspaced by small golden pieces. Attached to it was a picture of Sai Baba surrounded by a golden rosette frame about two inches in diameter. This necklace was presented to Mrs. Krystal.

4. *Appearance of Vibuti (Holy Ash)*

The fourth incident of possible materialization that we observed and will report upon here occurred in the open. We sat cross-legged on the ground in a long line of people as Sai Baba walked by. He stopped in front of Professor Hasra, a friend of Dr. Banerji, whom we mentioned above. Professor Hasra was sitting second to the left of K.O. and third from E.H. Sai Baba waved his right hand. As we were sitting on the ground and he was standing, his hand was slightly above the level of our eyes.

His palm was open and turned downwards, and his fingers were stretched out as he waved his hand in a few quick, small circles. As he

did this, we observed a gray substance appearing close to his palm. This substance appeared just below and at his palm, and Sai Baba seemed to grasp it into his fist with a quick downward movement of his hand as if to prevent it from falling to the ground. K.O., who sat slightly closer to Sai Baba than did E.H., observed that this material first appeared entirely in the form of granules, like very rough-grained sand. Sai Baba then poured the granules into the palms of Drs. Hasra and Banerji and most of them disintegrated into amorphous ash which they smeared on their foreheads. The point is that the granules were very fragile and would have lost their structure if produced by the magician's art of quick movements ("the hand is faster than the eye") which were invisible to us. When K.O. first saw the vibuti (holy ash), the granules were intact. This ostensible materialization of vibuti is a frequent occurrence and Sai Baba produces it several times as he walks among the crowd. We observed many such incidents, but only this one at so short a distance.

DISCUSSION

The alleged paranormal appearance and disappearance of objects has been a tough problem for psychical research in the sense that observations are rarely permitted under conditions which would exclude all possible normal causes. We were prepared to make instrumented observations with movie cameras and small sealed or locked enclosures wherein we hoped the objects would appear. Unfortunately, we were told not to use these in Sai Baba's interview room. We filmed him outdoors, waving his hand and producing holy ash, but not at close enough range for decisive analysis. All we have are observations made under semi-spontaneous conditions. Therefore, all our conclusions have to be extremely tentative.

Let us spell out some hypothetical normal explanations for the incidents we observed:

1. We might have been in altered states of consciousness, like mass hypnosis, and have responded to skillful suggestion techniques by "seeing" what was not there and overlooking actual, observable events. For example, the late Carl Vett (personal communication) explained his observations of the Indian rope trick in this way. We are both psychologists and can state with confidence that we did not undergo any altered states during our interviews with Sai Baba. We were very much on our guard at all times. Moreover, the objects produced (the double rudraksha and the gold ring with the enamel picture) are still in our possession.

2. The objects might have been provided by an accomplice in the interview room. This is not possible because objects also appeared when we were alone in the room with Sai Baba. Moreover, the seating

positions often excluded such a possibility, e.g., when he was seated at some distance from the other persons. Those present were visitors who varied from interview to interview. Only at our first visit in 1973 was an interpreter used who was also an "officebearer" of Sai Baba's organization.

3. The interview room might have contained concealed devices which somehow ejected the objects we observed. The room was barren of anything which could be so used. Sai Baba usually sat cross-legged on the concrete floor out of reach of any possible containers, such as a shopping bag on a windowsill, in which packages of vibuti or other small objects might be concealed. The place where he sat varied from interview to interview, and he was not positioned at one particular spot when the incidents occurred. He also produced objects outdoors and in a private room.

4. Sai Baba might have concealed the objects on his person and produced them by sleight-of-hand. We heard rumors about this possibility which suggested hiding places such as the sleeves of his robe, hidden pockets, and even his hair. However, we found no one who could offer firsthand observations or who could name someone who had made firsthand observations supporting this hypothesis.

We consider hypotheses 1-3 to be unreasonable and not worth further discussion. However, the sleight-of-hand hypothesis needs careful consideration because magicians do make objects seem to appear and disappear by this method.

Now back to our experiences with Sai Baba. We made some 20 observations of ostensibly paranormal appearances of objects in his hand. None of these occurred under controlled conditions and we were not able to examine him physically or to take other necessary precautions. Therefore, at this stage we obviously do not have sufficient grounds for accepting the claims made about the genuineness of the reported phenomena. It must also be stated that under the given conditions we were not able to detect any evidence of fraud.

Consideration of the following points leads us to regard Sai Baba's phenomena as possibly paranormal:

1. Lengthy history without clear detection of fraud. According to those who have had a long association with Sai Baba, the seemingly paranormal flow of objects has lasted for some 40 years, or since his childhood. Most of the persons we met who had had even just one meeting with him reported having observed some ostensible materialization phenomena. We did not meet anyone who claimed personal observations indicative of Sai Baba having produced the objects by normal means.

2. Reports of the occurrence of other psi phenomena, such as ESP over distance, giving messages in dreams, healing, out-of-body projections collectively perceived, and PK of heavy objects.

3. Variety of circumstances in which objects appear: during private interviews, while traveling in a car, outdoors in the presence of crowds, in private homes, etc. Almost every time we saw Sai Baba, in public or in private, objects were produced.

4. Production of objects apparently in response to a specific situation or on the direct demand of the visitor. We encountered many witnesses who testified as to such occurrences: the appearance of statuettes of a deity on request, a ring with the picture of a visitor's favorite deity, etc.

5. Reported production of large objects, e.g., a bowl the size of a dinner plate, and a basket of sweets 20 inches in diameter.

6. Production of objects at a distance from Sai Baba, such as prayer beads appearing on the windshield of a car being driven along an open country road, holy ash appearing on Sai Baba's pictures (observed by two senior research scientists), fruit appearing directly in the visitor's hand, etc.

7. Several prominent scientists in India have had the opportunity of observing Sai Baba extensively and have become convinced about the genuineness of the phenomena. Among them is Dr. S. Bhagawantam, former director of the All India Institute of Science (the most prestigious scientific institute in India) and a prominent nuclear physicist in India. We have already mentioned Dr. D. K. Banerji and Dr. P. K. Bhattacharya in the Department of Chemistry at the All India Institute of Science in Bangalore. We can also mention Dr. K. Venkatessan of the same institute who has worked in research at M.I.T. and Stanford University. Further we can mention Dr. V. K. Gokak, a former president of Bangalore University. We have met all these men, and they told us of a number of phenomena that they had observed in a variety of circumstances.

Of special interest, of course, is the wearing apparel in which objects might be concealed. Sai Baba wears a robe with sleeves of about the same length and width as those of our Western jackets. This garment is buttoned from the neck down, about the length of the sternum bone, and does not have any other openings or pockets that we could see; it is all one piece down to his ankles. As stated above, we had the opportunity of examining one of his discarded robes and found no pockets, slings, suspicious seam corners, or any other hiding places in it.

Another point brought out by several witnesses was that the kinds of things Sai Baba produces, for instance ash, would soil his robe if they were concealed in it. We saw him produce ash as many as four times during one appearance. After each production his hands were dirty with ash, but neither we nor the witnesses we interviewed ever detected ash on his clothes. Sometimes the ash was produced in large quantities—two palms held together filled with ash, as witnessed by

a college teacher and her psychologist husband. Among other things produced were Indian sweets and foods cooked in butter, which would certainly soil anything they contacted. One of us (E.H.) observed the production of oil, and a medical doctor who was formerly a faculty member at a medical school observed amrita, also a liquid, coming out of Sai Baba's hands.

We consulted a professional magician living in New York, Douglas Henning, who has given sophisticated performances in large cities in several countries. He was recommended to us as one of the most knowledgeable magicians in the world. He viewed a movie on Sai Baba and discussed our observations of objects appearing and disappearing. He was certain that he could by his magician's art duplicate all the cases he saw on the film. However, he considered the ring incident to be beyond the skills of magicians. He also said that if Sai Baba does produce objects upon demand, this would be a feat no magician could duplicate.

We also interviewed a British-trained Indian dental surgeon, Dr. Eruch Fanibunda, who is an amateur magician and has written two books on magic. He has taken motion pictures of Sai Baba and traveled with him, but has not observed any magician's tricks.

The most impressive incident we personally observed was the disappearance of the enamel picture of Sai Baba from K.O.'s ring. The sleight-of-hand hypothesis seems inapplicable because Sai Baba's hands, or those of potential accomplices, never came near the ring during the incident. We do not have a reasonable normal explanation for this disappearance.

We realize that without adequate experimental conditions the evidence will never be conclusive. However, it seems to us that the variety and richness of the phenomena associated with Sai Baba may provide unique research opportunities for both Western and Indian scientists.

REFERENCES

BENDER, H. New developments in poltergeist research. *Proceedings of the Parapsychological Association,* 1969, 6, 81-102.

CARRINGTON, H. *The Physical Phenomena of Spiritualism.* New York: Dodd, Mead, 1920.

CARRINGTON, H. *The American Seances with Eusapia Palladino.* New York: Garrett Publications, 1954.

CROOKES, W. *Researches in the Phenomena of Spiritualism.* London: Burns, 1874.

EISENBUD, J. *The World of Ted Serios.* New York: Morrow, 1967.

EISENBUD, J. The case of Florence Marryat. *Journal of the American Society for Psychical Research,* 1975, 69, 215-233.

HANNESSON, G. Remarkable phenomena in Iceland. *Journal of the American Society for Psychical Research,* 1924, 18, 239-272.

HOPE, C. Report of a series of sittings with Rudi Schneider. *Proceedings of the Society for Psychical Research,* 1933, 41, 255-330.

KASTURI, N. *Sathyam, Sivam, Sundaram: The Life of Bhagavan Sri Sathya Sai Baba.* Parts I-III. Bangalore, 1973-74.

MURPHET, H. *Sai Baba, Man of Miracles.* London: Frederick Muller, 1971.

OWEN, A. R. G. *Can We Explain the Poltergeist?* New York: Garrett Publications, 1964.

RICHET, C. *Thirty Years of Psychical Research.* (Trans. by S. de Brath.) New York: Macmillan, 1923.

ROLL, W. G. *The Poltergeist.* New York: New American Library, 1974.

SANDWEISS, S. H. *Sai Baba. The Holy Man and the Psychiatrist.* San Diego, California: Birth Day Publishing Company, 1975.

SCHRENCK-NOTZING, A. VON. *Phenomena of Materialization.* New York: Dutton, 1920.

SCHULMAN, A. *Baba.* New York: Viking Press, 1971.

YOGANANDA, P. *Autobiography of a Yogi.* Los Angeles: Self-Realization Fellowship, 1946.

Department of Psychology
University of Iceland
Reykjavik
Iceland

American Society for Psychical Research
5 West 73rd Street
New York, N.Y. 10023

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[11]

The Psychic Reading

RAY HYMAN
Department of Psychology
University of Oregon
Eugene, Oregon 97403

MIKE, A FREE-LANCE WRITER, visited a number of psychic readers to gather material for an article. He obtained a reading from each one. He also interviewed me to see how a psychologist would react to his experiences. Only one of the psychics, a palm-reader called Barbara, impressed Mike. In his article he wrote that, "This woman was studying lines in my hands and telling me, with devastating accuracy, about my strengths and weaknesses, my obsessions and my yearnings, my talents and my needs. As I drove away . . . I felt ready to chuck skepticism forever."

Mike discussed with me his experience with Barbara shortly after his visit with her and before he wrote his article. He had some difficulty in articulating just what it was about the reading that had so impressed him. He told me that during the first part of the session, Barbara made the usual assortment of general and ambiguous statements. His mind began to wander and he found himself thinking about a problem he was having with his girl friend. Suddenly his attention was brought back to what Barbara was saying. He heard her say, "I feel that you are worried about a relationship."

This conjunction in time between her statement about a relationship and his conscious concern about his relationship with a girl friend hit Mike with an emotional wallop. He had no doubt that the relationship Barbara was talking about was *the* relationship he was worried about. This emotional release converted an otherwise typical reading into something special. And Mike was now convinced that Barbara had some special powers.

From the skeptical viewpoint, Mike's account does not justify attributing special knowledge or insight to Barbara. Mike was impressed by her. To account for this impression he pointed to her accuracy in telling him he was worried about a relationship. We do not need to assume any powers to account for this sort of accuracy. We can safely assume that any individual who comes to a psychic has some sort of a problem with a relationship. Mike, of course, is sufficiently sophisticated to realize this. Yet he could not shake the conviction that Barbara's ac-

curacy was more than just the use of statements that could apply to anyone. He was sure that Barbara had somehow tapped into his innermost secrets. And he had experienced, as a result, a compelling and rewarding emotional experience.

We can raise a number of questions about Mike's encounter with the psychic reader. The question that naturally occurs to the psychologist involves validity. Did the psychic's statements actually correspond to the facts of Mike's personality and situation? Did they do so in a way that would differentiate him from other clients? Notice that this same focus on the accuracy of the reading was the basis that Mike employed to justify his positive evaluation of Barbara. Another question asks what is it that the reading actually does for the client. Why do clients such as Mike experience the reading as both revealing and helpful? Notice that a reading need not be accurate to be helpful. We can ask, also, how much the client gets out of the reading in relation to how much he or she puts into it.

This last question relates the psychic reading to the topic of this conference—The Clever Hans Phenomenon. When the horse, Clever Hans, was asked a question, he would often give the correct answer by tapping an appropriate number with his hoof.[1] This would occur under circumstances that seemingly precluded the horse's actions being under the control of signals from the owner. Because the questioner knew he was not cuing the horse, he assumed that Hans's answer was a response to the verbal question and that the answer, by being correct, revealed the conscious understanding of the question and the requisite knowledge to supply the answer. In fact, Hans was responding to a simple, involuntary postural adjustment by the questioner, which was his cue to start tapping, and an unconscious, almost imperceptible head movement, which was his cue to stop. The horse was simply a channel through which the information the questioner unwittingly put into the situation was fed back to the questioner. The fallacy involved treating the horse as the source of the message rather than as a channel through which the questioner's own message is reflected back.

The psychic reading shares this fallacy with the Clever Hans situation. In most such readings the psychic is simply a channel through which information unwittingly emitted by the client is fed back to the client. The client typically assumes that the message originates from some secret or occult source to which the psychic has access.

But the psychic reading is richer and more complicated than the Clever Hans situation. Both situations deal with an individual who is unknowingly both the source and the destination of the message. In both cases, the assumption is made that actual communication is taking place because the information being received is "accurate." Hans

answers the questions accurately. The psychic apparently tells the client things that he or she accepts as accurate. And this is just how it should be if actual communication were taking place. But it is just at this point in determining the accuracy of the communication that we notice an added feature that contributes to the success of the reading.

The questioner has no difficulties in deciding whether Hans has answered correctly or not. Either Hans has tapped the appropriate number of times or he has not. But the output from a psychic reading is both complex and highly ambiguous. The referent is the life history, personality, concerns, and problems of the client. And the language of personality description is notoriously difficult to interpret and apply.[2] This gives rise to what has become known as the "Barnum effect"—the phenomenon whereby people willingly accept personality interpretations comprised of vague statements with a high base rate occurence in the general population."[3]

Beginning with the 1949 publication by Forer on "the fallacy of personal validation,"[4] an area of research has emerged which follows more or less the same paradigm: (1) the subject completes a personality test or supplies information relevant to an assessment procedure; (2) the subject waits while the assessment information is processed; (3) the subject then receives a personality sketch allegedly derived from the assessment information; and (4) the subject rates the sketch for its "accuracy." In actual fact, all the subjects receive the same stock spiel composed of "Barnum" statements. The results are quite consistent and robust. When the subjects believe that the sketch was specifically meant for them they tend to rate it as a highly accurate and unique description of themselves.[3]

The appeal of these Barnum-type statements is so strong that it even overrides personality sketches especially written for the subjects. Sundberg, for example, administered the Minnesota Multiphasic Personality Inventory to 44 students.[5] Subsequently each student was presented with two personality sketches. One was written by a trained psychologist especially for the student on the basis of the student's answers to the inventory. The other was a fake sketch. Each student was asked to choose "Which interpretation describes you better." Of the 44 students, 26, or almost 60%, chose the fake sketch over the one written especially for them. Since Sundberg's study, several other investigators have demonstrated the same result.[3]

In a way this outcome should not be surprising. The fake sketches are composed of items that are true of almost everyone. The personalized sketches are composed of statements that discriminate the subject from others. When Barbara told Mike that he was worried about a relationship, we should not be surprised that Mike accepted the state-

ment as accurate because such a statement applies with high probability to just about every young, single male. But Mike accepted the statement not because he recognized its universal applicability, but rather because he understood it to apply to his unique circumstances. And subjects in the experiments on the Barnum effect accept the sketch as not only accurate, but as uniquely descriptive of themselves as distinct from others.

Much of the research on the fallacy of personal validation tries to isolate the conditions under which the fake sketch will be read and accepted as a unique description. Subjects *can*, under appropriate circumstances, recognize that the items in the sketch apply to them just because they are universally true for everyone. If the subject is handed the fake sketch and simply told it is a general description rather than one made especially for the subject, then he or she is less likely to accept it as a unique self-description. But the subject will not only accept as accurate the very same sketch, if presented under the belief that it was prepared especially for him or her, but will fail to realize that it is just as accurate for the general population. In addition, as a result of this acceptance, the subjects also increase their faith in the assessment procedure and the skill of assessor.[3]

Almost certainly the situations in which subjects accept the fake sketch as unique and those in which they recognize its general applicability are experienced quite differently. The meaning of the sketch, like the meaning of any literary product, cannot be separated from the reader or listener. As the semioticians and structuralists keep emphasizing, there is no unique relationship between signifier and signified—nor is there a unique referent or reading for any given message.[6] The reader of the sketch that is not allegedly prepared especially for him or her is doing something different from the reader of the sketch that is allegedly prepared especially for him or her. Each reader is actually reading a different sketch.

All cognition and all sign-interpretation (the two are almost synonymous) involve a heavy contribution by the recipient in addition to that of the sign and the sender. In both the psychic reading and the Clever Hans situation, the contributions of the receiver almost totally determine the message and its interpretation. The fallacy in these situations is due to the fact that the receivers do not realize how much of the message and its meaning is their own contribution.

The Barnum effect is sufficient to account for the apparent accuracy in certain kinds of psychic readings. In some types of readings, the client is presented with a complete sketch. In such situations the psychic or "sender" need not even be physically present (getting an astrological writeup through the mail, receiving a printout of personality

descriptions selected by a computer, etc.). No opportunity is afforded for the client to ask questions, clarify statements, or to agree or disagree while the message is being delivered. Nor does the psychic have the opportunity to alter or modify statements as a result of the client's reactions.

In contrast to such a static reading, most psychic readings are "dynamic" in the sense that the message is delivered to the client sequentially with the opportunity for the client to interact with the psychic during this process. It is in these dynamic readings that the Clever Hans effect combines with the Barnum effect to yield very effective results. In the typical reading, both the psychic and the client are equally victims of these two effects. Both believe that the client's acceptance of the reading and the client's conviction of having been helped stem from the psychic's access to hidden knowledge and to skills in diagnosing and advising.

But there is a class of psychic readers who are quite conscious of how they are managing the reading and the client to create the illusion that the psychic is the source of both the information and successful solutions to problems. These readers have sometimes written manuals to guide other readers. (Examples of such manuals, which I consulted, are listed in the bibliography.[7-13]) The manuals divide readings into two types: (1) the psychological reading and (2) the cold reading. The psychological reading, which corresponds to the static reading, involves delivering a stock spiel to the client. Because the sketch is usually memorized and delivered to several different clients, the studies of the Barnum effect apply directly. The cold reading is so called because the client is encountered without any prior knowledge. The cold reading employs the dynamics of the dyadic relationship between psychic and client to develop a sketch that is tailored to the client. The reader employs shrewd observation, nonverbal and verbal feedback from the client, and the client's active cooperation to create a description that the client is sure penetrates to the core of his or her psyche. The cold reader achieves this goal by feeding back to the client information the client has unwittingly revealed during the course of the reading. (Elsewhere I have written another version of how this is done and why it apparently works.)[14]

The manuals for both the psychological reading and the cold reading essentially provide guidelines for creating just those conditions that convince the client that the reading is accurate and that the accuracy is a function of the special knowledge of the psychic. These guidelines, while sometimes vague, clearly indicate that the writers operate under assumptions and theories very much in accord with modern cognitive science, literary structuralism, semiotics, and social cognition. None of

these writers, of course, are familiar with such disciplines. And their assumptions and theories antedate the development of these fields of inquiry.

But before I indicate some of the ways the manuals anticipate current academic disciplines that deal with communication and signs, it is worth considering again the issue of accuracy. The success of the psychic reading depends entirely upon acceptance by the clients—upon personal validation. And, on the surface, such acceptance seems to be related to how accurate the sketch appears to the client. But why should the client care or be concerned about such accuracy? The client accepts what the psychic says about him or her only if it agrees with what the client already knows. Why should the client pay for being told what he or she already knows?

Several answers come to mind. One is that many of the things that the client is being told are recognized or accepted as "true" but were not consciously considered previously. In this sense, the client is learning new things about himself or herself—but things that are acceptable because they "ring" true.

But probably the most important reason is that to the extent the psychic can tell the client things that are true, but which the client believes could not be known to the psychic through normal means, the psychic validates the belief that he does, indeed, have special powers. Like the shaman, the psychic reader is most successful if the client attributes to him mysterious powers and occult sources of knowledge.

The client does not patronize the psychic to discover things that the client already knows. The client is typically a person with a problem who seeks help. If the helper clearly has access to hidden knowledge and magical forces, this increases the chances that he can work wonders for the client. If the psychic can convince the client that he knows things that could only come from mysterious sources, then his powers are validated. So the client has a big stake in the accuracy of the psychic. And part of the success of the reading stems from the client's need to see the psychic as omniscient.

The manuals, then, can be seen as guidelines for creating in the client the illusion of the reader's omniscience. As such, they embody a theory or a set of presuppositions about how to manage the dyadic interaction so that the client "reads" the results in the desired manner. And these presuppositions, as previously indicated, seem to anticipate the ideas emerging from the variety of overlapping contemporary disciplines that study humans as sign-using systems. To illustrate this point, I will consider the types of advice provided by the manuals under four categories: (1) setting the stage; (2) preliminary observations and

categorization; (3) constructing the preliminary script; (4) delivering the message and revising the script.

Setting the Stage

The psychological and cold readers attribute much of their success to how the client is prepared prior to the reading. Such preparation consists of advertising, word-of-mouth accounts by previous clients, the preliminary introduction and statements to the client, the dress and mannerisms of the psychic, the furnishings and arrangement of the consultation room, and other ways that help to "define the situation" for the client and that specify what role the client is to play in this interaction.

Such preparations accomplish a number of important goals. The manuals specify that the psychic should make it clear from the outset that the psychic is the one who is in control of the situation. The psychic is not only experienced and an authority in this sort of interaction, but he or she has already proven that the psychic can do his or her part successfully. But, as the psychic will indicate to the client, the reading is a cooperative venture that requires the active participation of the client. This puts much of the burden for success on the client. It also emphasizes that the client has to collaborate with the reader to produce a satisfactory outcome.

It also reinforces the idea that the reader is omniscient and knows what he or she is saying and doing. If something that the reader later says does not tally with the client's beliefs or does not make sense, the client has been prepared to treat the apparent confusion as due to the client's own failure to adequately understand rather than to the psychic's lack of knowledge.

Some manuals further suggest ways to encourage this attitude in the client by having the reader state, as part of the preparation, something like, "During the course of the reading, impressions and images will come to me that will make no sense to me, but are very meaningful for your current or future situation. It will be up to you to make sense of these thoughts."

This setting of the stage recognizes what is now taken for granted in cognitive science—that the listener/reader reacts not to a message or test as such, but to the message as encoded. And this encoding must always be in terms of frameworks, concepts, expectations, and attitudes that the listener/reader brings to the situation. The role of the listener/reader in constructing the context and meaning of stories and literature is also recognized in semiotics and structuralism.[6] The concept of genre,

for example, is considered to be a guide to the reader on how to interpret the text. The same set of words, when set out in the form of a poem is "read" differently than when set out in the form of a brief note. And it is recognized that no matter how concrete and specific a message appears to be, it still can bear a variety of "readings."

The client who is actively processing the message from the psychic as a meaningful account of the client's current situation and the client who is skeptically processing the same message as a set of Barnum propositions are experiencing two qualitatively different types of communication.

Another very important part of the preparation of the client is not directly under the control of the reader. This is the client's objectives in participating in the reading. Most clients have problems, concerns, worries, and feelings of inadequacy. They seek solace, advice, support, or simply a good listener. For some the problems are chronic and of long standing. For others, the difficulty is a current and acute crisis. Such clients obviously have a vested interest in extracting the greatest benefit possible from their exchange with the reader. They are going to encode and react to what the reader says much differently than the casual curiosity seeker might. Even in the latter case, such as happened with Mike, the client who is merely curious or even skeptical is often surprised and impressed by what takes place.

But, given the highly involved client and the proper setting of the stage, even the crudest psychological reading—as evidenced by the work on the Barnum effect—is almost sure to meet with "success." But the readers have further items in their bag of tricks.

Preliminary Observations and Categorization

Even the manuals that supply the user with stock spiels to memorize and deliver suggest that some accommodations be made to take into consideration characteristics of the clients such as age, sex, socioeconomic status, and obvious signs of health. Sometimes such adjustments require minor alterations in statements indicating whether a problem was in the past, present, or future or adapting sexually and age-related references. In other cases, the manuals actually supply different stock spiels for different categories of clients: the young, unmarried female; the young, married female; the young, unmarried, male; the elderly woman; and so forth. Such adjustments based on preliminary observations and categorization of the client move the psychological reading towards the flexibility of the cold reading. The cold reader goes beyond this preliminary categorization and keeps refining and revising the reading to reach what is eventually a customized description.

All the manuals emphasize the surprising amount of specific information that one can pick up from a careful study of dress, physical characteristics, gestures and mannerisms, posture and attitudes, speech, jewelry, name, address, coat labels, and the like. The shrewd reader employs this information not only to make preliminary categorizations, but also to surprise the client with statements that the client believes could only have come from some occult source.

Constructing the Preliminary Script

The psychological reader, having set the stage and made some preliminary categorizations of the client, is ready to deliver the final version of the reading. This final version is a combination of facts based on the preliminary observations and a generalized or schematic script, which serves as a framework for ordering these facts. The psychological reading is based on the explicit assumption that there are universal features to all human lives as well as common characteristics and problems that face individuals in similar times and environments. Many manuals offer stock spiels for delivering a cradle-to-grave reading—with provisos for adjustments based upon sex, age, and other obvious characteristics of the client. Such a reading is based on the assumption that most of us live through the same major milestones: birth, childhood, school, work, marriage, children, and death. And during that life span we face more or less the same general problems such as career choice, love, sex, health, financial matters, and family.

The success of a book such as Gail Sheehy's *Passages* confirms what the psychological readers already know—that we share, despite our various lifestyles and unique histories, common problems and "predictable crises" at various ages such as the 20's, 30's, 40's, 50's, and beyond. The paperback edition says on its cover, "At last this is your story. You'll recognize yourself, your friends, and your lovers." It is just such recognition that the reader counts upon in his clients to make the reading succeed. Indeed, contemporary manuals strongly urge the psychic readers to study such books as *Passages* to better make their scripts realistic and convincing.

The use of a generalized, universal script around which to construct the reading contributes in many ways to its acceptability. Many of the ideas in structuralism and contemporary cognitive science testify to the value of such an underlying schema. Both the psychological and the cold reading are constructed on the basis that there are universal themes shared by all human lives. And this idea that diverse lives and unique histories are constructed from a limited set of constituent components motivates structuralist and semiotic approaches to literature,

myths, and cultures.[6] In psychology, the idea that we both encode and remember our experiences in terms of underlying schematas was emphasized by Bartlett in his classic *Remembering*.[16] However, it is only recently that psychologists and cognitive scientists have begun seriously applying Bartlett's insights into how we encode, understand, and remember prose.[17-19]

The reader's generalized script, among other things, acts both as a memory probe and an organizational framework for the client. To encode and make sense of what the reader is saying, the client has to supply the flesh to the skeleton. He retrieves from memory incidents and examples to instantiate the more general things being described. Mike did not simply hear Barbara say something about a relationship; he heard her talking about his particular problem with his girlfriend. Furthermore, this particular problem was now placed within an organized setting created by the underlying script and other relevant incidents that it brought to mind. Anything that Barbara said that did not make sense in terms of this script or anything that Mike thought about that was inconsistent or contradictory to the overall unfolding story would later be forgotten or difficult to retrieve. The script not only provides organization and meaning to the client's experiences, but it also strongly guides how and what he or she will recall from the reading. Both laboratory research and what we know about actual psychic readings predict that the client will remember mainly those things the psychic said that were consistent with the overall script and will also remember them in terms of those concrete memories that the client brought to bear in order to make sense of the reading.

Delivering the Message and Revising the Script

The cold reader does not stop with the preliminary script. He delivers the message in the presence of the client and modifies it as it unfolds on the basis of reactions and inputs from the client. This is where the Clever Hans phenomenon works to supplement the Barnum effect. The cold reader uses Barnum type statements organized around the preliminary script as trial probes. These statements are modified, withdrawn, revised, or elaborated in terms of the reactions of the client. These reactions are typically nonverbal such as pupillary enlargement, eye movements, postural adjustments, facial changes, and the like. The reader employs these cues to quickly pin down those topics of most interest to the client and to gauge when he is or is not on the right track.

Often the client emits verbal responses as well. These vary from exclamations to questions and comments. The manuals encourage the reader to promote such verbal feedback. One technique for doing this,

for example, is "fishing"—the psychic seemingly makes statements but phrases them as subtle questions. "I see two dark and tall men in you life—do you recognize them?" "I'm getting a vague image concerning a deed or some such financial document—does that make sense to you?" Once the client does begin responding with questions and comments, the reader takes pains to reward such behavior with careful and attentive listening. By listening to whatever the client says, the reader accomplishes many goals simultaneously. The client is convinced that the reader is sincerely interested. He or she probably has a strong desire for someone simply to listen. The attention encourages further such responses by the client. And, finally, the psychic is carefully storing this information in memory for later use in the reading.

When the client is not talking, the reader overwhelms him or her with a steady stream of patter. Part of the reason for this fluent outflow is to continually monitor the client's reactions and pick out topics and items that create the most reaction. Another reason is to prevent the client from realizing how much the client has already said. The successful reading, as the manuals make clear, is one in which the client ultimately is given what he or she considers a devastatingly accurate and penetrating personality analysis. The information for this analysis has come from both the verbal and nonverbal behavior of the client. The reader has merely fed back to the client information which the client has unwittingly supplied to the reader. The reading succeeds because the client is completely unaware that he or she has been the source of the description.

In effect the reader is like a ghost-writer who helps the client construct a coherent autobiography on the basis of information supplied by the client. In the typical psychic reading, both the psychic and the client falsely attribute the accuracy of the final reading to the occult sources to which the reader has access. In the cold reading, the reader is fully conscious that the client is the source of all the information and takes steps to maximize the client's contributions.

The impact and success of the psychic reading go beyond the illusion of accuracy. The client's acceptance is ultimately based on an emotional experience rather than a dispassionate assessment of the accuracy. Indeed, as the case of Mike and Barbara suggests, the attribution of uncanny accuracy to the reader probably stems from the emotional impact achieved during the reading. The attribution of accuracy is based upon information that the client has supplied to the reader. In a sense, the reading which the psychic later gives back to the client contains nothing "new." In another sense, however, the repacking of this material and putting it into a coherent order provides new insights to the client. In addition, the client is now looking at his or her memories and ex-

periences from a new vantage point. This provides something similar to what Shklovsky attributes to the function of literacy devices. They serve not to represent familiar events, but rather to make them strange—to defamiliarize them.[6] In a way, the reading does for the client's self-concept what Poincaré claims that mathematical discoveries do for already familiar concepts—"they reveal to us unsuspected kinship between other facts, long known, but wrongly believed to be strangers to one another."[20]

It is thus conceivable, then, that despite the false attribution of secret knowledge to the reader, the client emerges from the reading with a new and more adapative model of his situation. He or she may have a new insight into the conflicts and problems that precipitated the consultation. And new alternatives for coping with the situation may have been opened up. Whether or not the new perspective and vision of the client's life and situation is ultimately beneficial, it is easy to understand how the client might experience the reading as revealing and rewarding.

Much more can be said about the psychic reading. But I hope I have said enough to convince you that it is a rich and challenging opportunity for studying the ways signs, witting and unwitting, combine to create illusions of communication in dyadic settings. Both the Barnum effect and the Clever Hans phenomenon combine to induce the overwhelming conviction that the psychic is the source rather than a mirror of an accurate appraisal of the client and his or her circumstances. In some cases it might be revealing to look upon the reader as a catalyst that aids the client in the construction of a coherent and helpful self-description. In general, however, the psychic reading, like the Clever Hans case, is but one of many illustrations of the human propensity to project meaning into situations even when the situations, themselves, have no meaning.

References

1. Pfungst, O. 1965. Clever Hans. Holt, Rinehart & Winston, New York, N.Y.
2. Bromley, D. B. 1977. Personality Description in Ordinary Language. Wiley & Sons, New York, N.Y.
3. Synder, C. R., R. J. Shenkel & C. R. Lowery. 1977. Acceptance of personality interpretations: The "Barnum Effect" and beyond. J. Consult. Clin. Psychol. 45: 104–114.
4. Forer, B. R. 1949. The fallacy of personal validation: A classroom demonstration of gullibility. J. Abnormal Soc. Psychol. 44: 118–123.
5. Sundberg, N. D. 1955. The acceptability of "fake" versus "bona fide" personality test interpretations. J. Abnormal Soc. Psychol. 50: 145–147.
6. Hawkes, T. 1977. Structuralism and Semiotics. University of California Press, Berkeley, Calif.

7. ANONYMOUS. 1971. Pages from a Medium's Notebook. Micky Hades, Calgary, Alberta, Canada.
8. BOARDE, C. L. 1947. Mainly Mental: Volume I: Billet Reading. Globe Service, New York, N.Y.
9. CORINDA. 1968. Thirteen Steps to Mentalism. Louis Tannen, New York, N.Y.
10. HESTER, R. & W. HUDSON. 1977. Psychic Character Analysis: The Technique of Cold Reading Updated. Magic Media Ltd., Baltimore, Md.
11. MAGNUSON, W. G. 1935. The Twentieth Century Mindreading Act or the Modern Spiritualist Medium's Act. Albino, Chicago, Ill.
12. NELSON, R. A. 1951. The Art of Cold Reading. Nelson Enterprises, Columbus, Ohio.
13. RUTHCHILD, M. 1978. Cashing in on the Psychic. Lee Jacobs Productions, Pomeroy, Ohio.
14. HYMAN, R. 1977. The Zetetic 1 (Spring/Summer): 18–37.
15. SHEEHY, G. 1977. Passages: Predictable Crises of Adult Life. Bantam Books, New York, N.Y.
16. BARTLETT, F. C. 1932. Remembering. Cambridge Univ. Press, London, England.
17. BLACK, J. B. & G. H. BOWER. 1980. Story understanding as problem-solving. Poetics. In press.
18. BOWER, G. H. 1976. Experiments on story understanding and recall. Q. J. Exp. Psychol. 28: 511–534.
19. BOWER, G. H. 1978. Experiments on story comprehension and recall. Discourse Proc. 1: 211–231.
20. POINCARÉ, H. 1955. Mathematical creation. *In* The Creative Process—A Symposium. B. Ghiselin, Ed. pp. 33–42. Mentor, New York, N.Y.

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[12]

An Overview of Quantitatively Evaluated Studies With Mediums and Psychics

SYBO A. SCHOUTEN

ABSTRACT. From the beginning of scientific research in parapsychology in the 1880s, mediums and psychics have generated much interest. However, no means were then available to make objective assessments of the verbal impressions of psychics and mediums. Interpretation of the data was purely a subjective matter. During the 1930s, the first attempts were made to develop methods for quantitative, and later on statistical, evaluation of purportedly paranormally-obtained statements of mediums. The history of the development of these methods is primarily a matter of the discovery and elimination of possible sources of errors. As a result, today satisfactory methods are available. A very condensed overview is presented of all studies in which statements of mediums or psychics have been quantitatively evaluated. The main question asked in most of these studies was whether a significant number of correct statements deviated significantly from chance expectation. Another question, less often addressed, was whether psi ability is necessary to explain the correct statements. The present study indicates that the number of studies with significant positive results is rather small. Moreover, in most of these, one or more potential sources of error were present that might have influenced the outcome. It seems, therefore, that there is little reason to expect psychics to make correct statements about matters unknown at the time more often than would be expected by chance. An explanation is offered for the apparent successes of psychics in everyday-life consultations that takes into consideration the role of the client or sitter, which is generally underestimated. Extremely put, it is proposed that it is the client who makes the psychic. Nonetheless, under certain conditions, which are described, asking advice from a psychic can be meaningful.

Although spontaneous paranormal experiences are frequently reported, few people claim to be able to have such experiences more or less on command. Such persons, known as mediums or psychics, have always drawn much attention. Some of them apply this presumably paranormal ability professionally to provide people who consult them with information on matters unknown at that moment, such as the fate and whereabouts of a missing person. A common procedure for psychics is to obtain seeming paranormal impressions with the help of an inductor (or token object), such as an item belonging to the target person, about whom information is requested. Known as object-reading today, an older term for this method was psychometry. However, the use of an inductor is by no means a necessary requirement. Many psychics provide impressions about target persons without the use of any prop.

The fact that psychics openly apply their paranormal abilities to advise others and that they sufficiently impress others with these abilities to attract more clients and to build a reputation for themselves distinguishes them

from other persons who also claim or believe themselves to be paranormally gifted, among them the self-deluded and outright frauds, but who fail to convince others of their "gift." Thus, psychics are persons who impress others by their achievements, and that provides the reason to study whether their claim that their success is due to some paranormal faculty is justified. As can be expected, the research discussed below has mainly been carried out with "stars," that is, psychics who had a great reputation in their day.

I use the terms *mediums* and *psychics* in this paper, and as regards the result of their activities, which is to provide information not normally known at the time, there is no difference between them. However, mediums and psychics do differ as regards their belief as to the origin of the information they receive. Mediums believe that they are able to communicate with the deceased and that the deceased are the source of the information they receive, often while in trance. Psychics think in terms of a paranormal ability, which is possessed by a few persons but not by most. Mediumistic beliefs became widespread in the last century with the rise of Spiritualism. Psychics can be considered as a more modern variant of mediums, and thus is the term most often used in this paper.

Despite the fact that mediums and psychics were in the center of interest from the beginning of psychical research in the 1880s, the possibilities for meaningful scientific research with them have been limited by the many difficulties associated with studies involving verbal material, or free statements. This may be one reason why after over 100 years of research in parapsychology there is still no clear answer to the question of what significance should be attributed to the activities of psychics. This is not only a matter of academic interest. Many psychics are active today, and frequently people want to know from parapsychologists whether they should or should not consult such persons, and if they do, what they might expect and how they might evaluate the statements made to them by the psychic or medium. As scientists in this field we have to provide answers to such questions. This article is an attempt to provide these answers, based on the quantitative and qualitative research data available at present.

History of Investigations of Verbal Statements of Mediums and Psychics

The first extensive studies of verbal statements of mediums appeared about 100 years ago in the publications of the British and American psychical research societies. These studies were purely descriptive. Hundreds of pages were devoted to transcripts of readings of mediums and discussions of interpretations and the validity of the mediums' statements. Later the interest decreased, partly because the possibility of gaining new knowledge from this material seemed to be exhausted. Another contributing factor might have been the development of more sophisticated research methods in science. The subjective estimation of the significance of data

became less acceptable and was gradually replaced by the application of quantitative and statistical evaluations.

In experimental research, the work of J. B. Rhine (1934a) in card-guessing, initiated in the 1920s because it was amenable to statistical evaluation, was a clear shift in this direction. His first research with the well-known mental medium Eileen Garrett involved not readings but card-guessing experiments (Rhine, 1934b). On average, Garrett obtained very significant scores during three weeks of testing, especially in the telepathy condition. She did not perform differently on ESP tests when she was in trance and under the control of the spirit "Uvani" than when she was in her normal waking state, which favors the hypothesis of a common capacity serving both personality states. Rhine concluded: "The abilities found in Mrs. Garrett's two states were within the range of those found in the eight normal young student subjects with whom my assistants and I have earlier worked" (Rhine, 1934b, p. 110).

At the end of the same decade, the first steps were taken to develop methods for the quantitative evaluation of verbal statements of mediums. Most of these methods were primarily, but not exclusively, restricted to verbal material. Other forms of communication, such as drawings, could also be evaluated. Progress in this area has been slow compared to developments in other areas of parapsychological research, primarily because of the difficulties involved in the evaluation of spontaneous verbal and other material. In addition, the amount of data and time involved in studies with mediums and psychics is considerably higher than in most experimental studies. Nonetheless, sufficient research involving quantitative analyses of statements of psychics has been carried out to allow us to state that today reasonably adequate evaluation methods are available and that the data collected with these methods enable us to draw some definite conclusions.

As Bender (1957) rightly stated, a quantitative analysis provides only a limited view of the data. A qualitative analysis, that is, an analysis of the material based on a subjective interpretation in psychological terms, is much more meaningful. A qualitative analysis can show, for instance, that ostensibly unrelated statements have a strong relationship because they all relate to some strong emotional experience in the life of the target person. Such a relationship can hardly be evaluated in a statistical sense. In addition, a statistically significant effect without a meaningful interpretation is rather useless. For this reason, Bender considers that quantitative analyses have mainly a supportive function. First it must be demonstrated quantitatively that the statements of the psychics can be considered as especially applicable to the target person. Only then is it worth taking the next step, which is to interpret the statements in qualitative and psychological terms.

The aim of this paper is to review the primary quantitatively evaluated studies and to assess the value that can be attributed to the statements of mediums and psychics. Many people seem to have a high opinion of the abilities of psychics. The reason for this may be that psychics actually do have an ESP ability that they apply in consultations. Most of the studies

reviewed here tried to answer the question: Are the statements of psychics about the target person, who is unknown to the psychic, more often correct than can be expected by chance? Less often a related question is asked: If this is the case, is it necessary to attribute a paranormal explanation to this result? After all, the real issue is whether these data can be explained paranormally or whether familiar psychological processes will suffice. This overview may not include *all* relevant studies, but I believe that the material is sufficiently representative to allow for conclusions to be drawn. In the presentation of the material a more or less chronological order is maintained. In view of the large amount of data, the discussion of individual studies is limited to only the most important data. In view of the number of studies, the large diversity in material and analyses applied, and the differences between the various studies in sources of error that may have affected the outcomes, no meta-analysis was carried out.

The Investigations

The First Attempt by Hyslop

After James H. Hyslop of the American Society for Psychical Research (ASPR) published his first report on his experiments with the trance medium Leonora Piper, Hyslop (1919, p. 5) reports that Podmore publicly declared that nearly all the facts could be explained by shrewd guessing and inference, together with chance coincidence. According to Hyslop, there was not the slightest excuse for Podmore's hypothesis, except the perversity of an intellect secure in "ad populum" methods. Nevertheless, Hyslop (1919) felt obliged to add some more weight to this opinion by demonstrating that coincidence could not account for the results. He started with a lengthy discussion on the degree of (im)probability that Mrs. Piper could mention so many names that he counted as applicable to his situation (Hyslop himself was the target person). But then he described his main control study, an attempt to arrive at a probability value for the scoring rate of the entire session. The session statements were rephrased in the form of 105 questions. These were then sent to 1,500 persons, of whom 420 returned the questionnaire. Based on the number of affirmative replies, Hyslop calculated the probability of the correctness of each statement. Thus, if 42 persons (of the 420) rated a question as applicable to them, the probability that that statement is true would be 1 in 10. These 105 probabilities were then multiplied with each other to obtain a final value for the entire session because, as he noted, he himself could answer every single question in the affirmative. But the justification for multiplying the probabilities depends not only on the correctness of the statements. It also assumes that the statements and affirmations are strictly independent. For instance the two statements "the person is married" and "the person has children" are, certainly in Hyslop's day, two nonindependent and related statements. If the one is correct, the likelihood of the other being correct

is greatly increased. As could be expected, Hyslop arrived at an astonishingly low p-value, proving his point. However, this was not only the first but also the last study in which apparently the psychic obtained a 100% hitting rate.

The Saltmarsh Study

An extensive study, which lasted over a year, was carried out by Saltmarsh (1929) with the medium Mrs. Warren Elliott. She gave readings about target persons unknown to her. In 53 of the sessions, the target person was present but not identified; in 89 sessions the target person was absent. Extensive precautions were taken to exclude normal sensory information. Although usually the medium would enter a trance state for obtaining paranormal impressions, in sessions in which no target person was present it frequently happened that she obtained her impressions in her normal waking state.

The aim of the investigation was to test different classes of theories: (a) explanations based on the probability that the results were due to chance; (b) explanations based on ESP; and (c) explanations based on the survival hypothesis, that is, that the information was communicated by deceased persons. It turned out that the sessions in which the target person was present appeared to be much more successful than the sessions in which no target person was present. Saltmarsh was also impressed by how stereotyped the medium often was in her description of the target persons: specific themes like leg wounds and certain names were mentioned repeatedly.

In order to test his first theory, Saltmarsh had to carry out some sort of quantitative analysis of the data. To this end, he introduced the concept of control studies. In a control study, the statements are not only judged on correctness by the target person for whom they were intended but also by control subjects. This method can show whether the statements are more applicable to the target person than to any other person, provided the control and target persons are comparable in important respects. For instance, it would not be appropriate to use male controls when the statements concerned a female target person.

Saltmarsh ran two control studies for sessions in which no target person was present. For the first, he selected a number of related sessions in which the statements were supposed to concern a young pilot who was killed in an air battle. The controls were six persons who had lost a relative under the same circumstances (as a young pilot in an air battle). For the second control study, statements from 15 successful, unrelated sessions were selected and given to 15 control persons to judge as to their applicability to their own situations. First, Saltmarsh classified the statements into three weighted categories: cliches ($w = 1$); clear statements ($w = 5$); and characteristic statements ($w = 20$). Based on the comments given by both target and control persons, he judged each statement as either correct (in

which case he added the weight of the statement) or as incorrect (in which case the weight was subtracted). Thus, total scores were obtained for each session for both the target person and the controls.

It appeared that in the first control study (involving the pilots), the score obtained for the target person was, on average, eight times higher than the scores of the control. In the second control study, the comparable value was 6.6. Saltmarsh considered this an indication that the medium possessed information about the target persons that she could not have obtained by normal means. Hence, he rejected the first theory. He also rejected the survival hypothesis, mainly because he felt there was no need to assume an active role on the part of the deceased to explain his results. He concluded, therefore, that ESP seemed to be the best explanation for Mrs. Elliott's performance. The scores appeared approximately two times higher for sittings in which the target person was present compared to sittings with no target person present. No difference in scoring was observed between sittings in which the medium went into trance compared with sittings in which the reading was made with the medium in a normal waking psychological state.

The method Saltmarsh used had some weaknesses, of which he was aware. One is that he himself was the final judge of the correctness of the statements, which might have introduced bias. But according to Saltmarsh, the judging task involved such a tremendous amount of work that he could not ask others to do it. Another weakness is the lack of a statistical assessment of the results.

The First Statistical Evaluation: The Saltmarsh-Soal Method

Two years later, Saltmarsh and Soal (1930) presented a method to statistically evaluate the verbal statements of mediums. The method was based on two criteria: (a) is the statement correct?, and (b) for what part of the population does the statement hold? In other words, what is the probability of the statement being correct? With advice from statistician R. A. Fisher, a formula was presented that combined the results of all correct and incorrect statements and their respective probabilities into one total score with an associated standard error. Divided, the two yield a standard-normal score (z score) with an associated probability of significance. In the second part of the paper, Saltmarsh applied the method to a session with Mrs. Elliott. The formula yielded the highly significant z score of 6.2, a result that impressed Saltmarsh so much that he wrote: "I submit that this result is such that the hypothesis that chance alone could have produced this amount of veridicality is definitely excluded" (Saltmarsh & Soal, 1930, p. 271).

In order to establish for each statement the proportion of the population for which it is true, Saltmarsh suggested employing committees of investigators who would be provided with the statements of the mediums and the comments on these statements made by the target persons. The latter

insures that the interpretation of the statement is the same for target person and committee. A disadvantage of this method is that it gave the committee members knowledge of the degree of correctness of the statement. The best solution, perhaps, would be for a third party to process the material first, and if necessary, add some clarification to a statement as to its proper interpretation before the committee started its work.

In the above-mentioned evaluation, Saltmarsh himself established the probabilities of correctness for each statement. Because he knew from the target person's comments which statements were correct, bias may have occurred. That judgments even by experienced persons can be extremely unreliable was pointed out by Findlater (1932), who described an American study that demonstrated that when seven experienced teachers independently rated the same 48 papers, the number of D grades varied from 2 to 37. Another problem Saltmarsh did not consider was the possible interdependence of statements. The probability of statements can be influenced by the correctness or falseness of previous statements. These factors may have influenced Saltmarsh's results. Another disadvantage of the Saltmarsh/Soal method is that it is unsuitable for application in sessions in which the sitter provides feedback. Such feedback often contains information that influences the probability of correctness of subsequent statements. However, this objection applies to nearly all methods developed later on to evaluate verbal material.

The First American Study

In 1933, J. F. Thomas obtained his Ph.D. from Duke University based on a study in which he presented a quantitative evaluation of mediumistic statements. The aim of the study was "the analysis of the mental content of mediumistic phenomena, in order that the explanation of chance hits may be more fully evaluated" (Thomas, 1937, p. ix). He acted as the target person in many sessions, anonymously visiting different mediums or having associates serve as sitters, such as his secretary. His data consist of 24 sessions obtained from the British medium Gladys Osborne Leonard and over 500 sessions with other mediums. The analyses presented in his book mainly concern the Leonard sessions. The statements made in all sessions were split into "topics" (statements about the same subject) and "points" (statements of a single fact). Three groups of sessions were distinguished: those he attended himself, those in which the records were taken by a secretary, and those that were taken by an uninformed intermediary. Thomas evaluated the correctness of the statements. For all groups of sessions he obtained a success rate of over 90% correct statements.

For the evaluation he adopted the two methods described by Saltmarsh (Saltmarsh & Soal, 1930). In one, 64 persons acted as a control group and scored 643 statements from nine readings for applicability to themselves. The controls were in various respects comparable to himself. From these

data, he acquired frequency distributions for the total sessions scores and thus established the significance of his own scores. The *z* scores he obtained were highly significant. The other method involved the establishment of probabilities for 16 statements from one session by asking two committees, working independently of each other, to provide these values. Based on these probabilities he applied the Saltmarsh/Soal (1930) formula. Again, he obtained highly significant results.

It is clear that Thomas himself played a major role in the evaluation and judgment process. He was the target person, sometimes the recorder who wrote down the statements of the medium, and the judge who eventually decided whether the statements were true. In addition, he was the one who selected the readings and statements that were used for the control studies. Very likely the medium, Mrs. Leonard, came to know quite a lot about Thomas in the course of the investigation. It would be of interest to know whether the success rate increased in the course of it. Another problem with this study is that the controls knew that they were controls, and as was shown by later studies (Bender, 1957), this was likely to bias their judgments in the direction of lower scores.

Pratt's Study With Mrs. Garrett

Two years after the publication of J. B. Rhine's (1934a) famous book *Extra-Sensory Perception,* which describes the results of his quantitative card-guessing ESP studies, J. G. Pratt (1936) published his quantitative evaluation of sessions with the medium Eileen Garrett. In his introduction to a reprint of this paper, Pratt (1960) notes that the two types of research, Rhine's and his own, had much in common despite their apparent differences: "They were both experimental studies done at Duke University and they were publications resulting from the new effort being made there to bring parapsychology into the mainstream of the strong movement toward quantitative experimental methods" (Pratt, 1969, p. 9).

Pratt began his paper with a discussion of the methods used to date for the evaluation of verbal material of mediums, and he states that despite the fact that progress had been made, a number of weaknesses were still apparent. The most important ones, in his opinion, were (a) that the people involved who rated the statements (researchers, target persons, controls) knew for whom the statements were intended, which could bias their judgments; and (b) the unreliability of estimating probabilities for the individual statements. As to the first problem, he states that although he believes this to be a source of error, he does not have any research data to support the belief. (We know now that his supposition was right.)

To eliminate these two weaknesses, Pratt introduced a different method. He carried out two series of sessions with Garrett involving 12 and 15 target persons, respectively. In Series I, the subject was present when the medium made her statements, whereas in Series II, the subject remained in an adjacent room where he or she was not able to hear Garrett's comments.

Series I was evaluated in the traditional way with the help of 25 control subjects, and it yielded a significant $z = 3.22; p = .001$. However, a new method of evaluation was applied in Series II. After the sessions were completed, Pratt, who had kept himself blind to the relationship between sessions and target persons, divided the verbal material into separate statements and then asked all target persons to evaluate each statement as to its applicability to their own situation. Hence, the persons who judged the statements were all blind as to whether the individual statements were intended for them or not. He then used the responses of the other 14 persons as control data for each session, from which a mean standard deviation for each session's scores was derived. The outcome of this series was even more impressive (a combined $z = 5.10, p = \ll.01$). However, this result is mainly due to two very successful sessions.

Pratt also checked for another possible source of error: the fact that persons might differ in their willingness to give affirmative responses. This could result in a significant positive session, as in the case where one subject might tend to rate most statements (intended for him and for the others) correct and the others not. Pratt concluded that this source of error could not explain his positive findings. Thus, even this methodologically improved method indicated that Garrett was able to provide more correct information on target persons than could be expected by chance.

Further Research in Britain

Unfortunately, Pratt's positive results were not confirmed by a replication study with Eileen Garrett conducted by Herbert (1937b). Earlier he had published a study (Herbert, 1937a) with a medium in which he investigated the extent to which target persons recognize the names mentioned in the medium's statements. The results were not significant. Half the names were recognized, the other half were not. Herbert's (1937a) study with Mrs. Garrett involved six sessions that were held under the conditions of Pratt's Series II, with the exception that Garrett did not meet the target persons at all. The target persons were all experienced sitters who, according to Herbert, had been successful with other mediums. Applying Pratt's method of evaluation yielded a nonsignificant overall result. Thirty control subjects rated approximately 600 statements on applicability to themselves. However, the Saltmarsh/Soal evaluation only confirmed the earlier finding: there was no trace of ESP. (It should be noted that Eileen Garrett typically had extrachance success with novel procedures but not with replications.)

A more successful study was conducted by C. Drayton Thomas (1939), but he may have been unacquainted with Pratt's and Herbert's work because he made no attempt to introduce any objective control on the value of his data. He asked a medium, the well-known Gladys Osborne Leonard, to provide information on the deceased father of the wife of a friend of his, and on the first, also deceased, wife of the same friend. Thomas claimed

that for the former about 75% of the statements proved to be correct; the percentage was about 51% for the latter. He attributes this difference to the possibility that the deceased woman was a less effective "communicator," and he rejected a possible influence of the fact that the first target person was a well-known public figure, whereas little was publicly known about the woman.

The Work of Hettinger

Also in the second half of the 1930s, an extensive study was carried out by John Hettinger, who was associated with London University. The investigation, which yielded him a Ph.D., is described in two books (Hettinger, 1940, 1941). Years later, he ran a replication study in collaboration with the ASPR (Hettinger, 1947). In 1949, Hettinger's work was seriously criticized by Christopher Scott (1949), who also presented data relevant to the judgment of the methodology Hettinger applied.

Hettinger's first investigation consisted of 12 series of inductor experiments in which two mediums participated. His friends and acquaintances were invited to submit the objects that served as inductors. Only Hettinger was present when the statements were made by the medium, and he served as the recorder. The 12 series differed mainly in the way he treated the data. In fact, the whole series was a search for a satisfactory evaluation method. Hettinger soon discovered that the subjects' judgments were strongly influenced by their knowledge that the statements were or were not intended for them. Instead of control persons, Hettinger introduced the notion of *control statements*. For instance, in Series IV, each statement was coupled with a statement randomly selected by Hettinger from a previous sitting of that medium and involving a different target person. The target person was asked to select the statement that he or she felt was most applicable to his or her situation. Another variant was that the statements from two sessions were randomly mixed and the two target persons were asked to select the ones they believed were intended for them. Apparently the latter method was abandoned because it gave rise to too many quarrels between the subjects. Another method provided two target persons with the two protocols of statements intended for them, and they were then asked to match each protocol with the right person. The originality of Hettinger's approach is that he avoided having the evaluation depend upon unreliable estimates of the probability of each statement being correct. He reduced the evaluation to a matching task, for which in principle the probabilities are known (e.g., 1 in 2 for the methods described above).

The main series involved over 600 sessions and over 6,500 statements. According to Hettinger, the overall result yielded a highly significant $z = 19.1$. However, after a thorough analysis of Hettinger's work, Scott (1949) concluded that Hettinger made a number of statistical errors and that the real significance level was between $z = 5.3$ and $z = 6.1$ (the associated

p values are still very low: $p \ll .01$). More seriously, Scott assumed that Hettinger made a number of methodological errors. Scott could not verify this, however, because Hettinger's publications are in many respects incomplete. According to Scott, an important source of error was the way Hettinger selected his control statements. In order to insure a probability of 1 in 2 for the target person to select the statement intended for him or her, it is vital that real and control statements be in every respect comparable, except that one was intended to be applicable to the target person and the other not. For various reasons, this comparability can be violated, and the real statement might obtain a higher probability of being selected (e.g., where the control statement is more specific than the real statement; the real statements show coherence and the controls do not; the real statements refer to the inductor whereas the controls do not; the person saw the control statement on a previous occasion; the control statements were selected by someone who knew the target person). One of the problems with Hettinger's study is that he, knowing all the relevant details, always selected the controls himself. Another objection is that Hettinger attended every session with the mediums and recorded all statements. In most cases, he probably knew the person for whom the statements were intended. It is not documented how far he was able to control himself and was able not to react to correct and false statements. It is also likely that occasionally he recorded in the form of abbreviated sentences, or did not write down statements he did not deem of sufficient interest, or gave an interpretation to statements when working out the notes that was colored by his knowledge about the target person involved.

Hettinger's (1941) second investigation involved an entirely different methodology. The target person bought a newspaper or magazine and read it at a prearranged time, making notes about the exact time he or she spent on each page or picture. At the same time, the medium described his or her impressions of the target person. The statements of the medium were then compared with the content of what the subject read or looked at during that moment. According to Hettinger, a margin of one minute in both directions was allowed in the comparison. In addition, he checked on whether the statements applied to the personal circumstances of the target person.

A number of subsequent experiments were variations on this theme. For instance, instead of newspapers and magazines 30 pictures were looked at one by one for two minutes. The main experiment involved 3,513 statements, of which, according to Hettinger, 585 appeared to be related to the target material and 322 to the target person. His second book, *Exploring the Ultra-Perceptive Faculty* (Hettinger, 1941), consisted mainly of examples of the correspondences observed. However, a statistical analysis of these data is not possible. Another objective is that Hettinger himself did all the judging, comparing the statements with the content of the paper or picture. It is clear that the rich content of this target material offers many opportunities to find such correspondences without recourse to ESP.

A Control Study Together With the ASPR

Even so, the results of Hettinger's (1941) second study impressed the ASPR sufficiently for them to repeat it, but this time it was set up in such a way that the data could be statistically evaluated. The first experiment followed the familiar design. However, as a control, Hettinger also compared the statements with another issue of the paper used in the session. A significant difference was observed in favor of the real target material. This result was not entirely convincing, however, because again all judgments were made by Hettinger, who knew what the targets and the controls were. To eliminate this objection, a second study was carried out in which two picture books of 60 pictures each were used. One book was selected as the target, and the subject looked at each picture in that book for one minute. Hettinger, who in this experiment did *not* know which book was used as the target, found 264 statements of the medium applicable to Book 1 and 362 to Book 2. He correctly concluded that Book 2 must have been the target. Although it thus appeared that the medium picked up information from the person looking at the pictures of Book 2, Laura Dale (in a personal communication to Hettinger dated October 13, 1946 cited in Hettinger, 1947, p. 108) suggested an alternative explanation, which was that Hettinger himself somehow might know that Book 2 was used as the target and became biased by that knowledge in favor of finding correspondences between the statements and that book. She urged Hettinger to have an independent judge redo the assessment. Hettinger complied with this request, and the results were dramatic. The independent judge found 126 statements corresponding to Book 1 and only 98 corresponding to Book 2. Not only the direction of the difference was reversed, but the total number of correspondences observed was less than half of those found by Hettinger. This clearly demonstrates how unreliable such judgments are.

A Control Study by Christopher Scott

It might well be that more value should be attached to Hettinger's judgments in view of his extensive experience in judging verbal material. In order to investigate this, Scott (1972) convinced Hettinger to participate in another control study. Scott acted as target person and looked at a picture book made up of 60 pictures. However, after the session Scott secretly replaced 30 target pictures with control pictures. Hettinger, believing that all 60 pictures had been targets, did the judging. Two such studies were carried out. According to Hettinger, the correspondences he found were comparable to the results of previous studies. It turned out that in the two studies, no difference was observed in the number of correspondences for target and control pictures. Although this does not prove convincingly that the findings of Hettinger's previous studies were due to bias in his judging of the material, it does demonstrate how easy it is to find

correspondences between statements of mediums and unrelated (control) pictures.

This finding was soon corroborated by the results of a study by Parsons (1949). Parsons sent the protocol of a session with a medium, which concerned a deceased person, to four persons who were comparable to the deceased in age and sex but who were still alive. These subjects rated the protocol in the belief that the reading had been intended for them. Two of the four subjects rated the protocol as very applicable to themselves; for one, the result was even better than it was for the real target person. Parsons concluded that protocols should not be judged only on the comments of the target person, as was the rule in all the older research with mediums.

The Pratt-Birge Matching Method

That this requirement was already recognized on the other side of the Atlantic is apparent from Pratt's first article on the subject (Pratt, 1936). His efforts concentrated on the development of better methods of statistical evaluation and eventuated in the *Pratt-Birge method* (Pratt, 1969; Pratt & Birge, 1948). The essence of the method is that a medium makes statements about a number of target persons, and each target person judges all statements on applicability to his or her own situation without knowing for whom each statement was intended. The statistical analysis is based on these responses and treats the data as a matching problem. If the protocols are matched with the target persons in correct order, a total score is obtained based on the number of correct statements. However, applying a different order yields a different total score, because in that case some or all target persons have rated protocols intended for other target persons. In the case of five target persons, and hence five protocols, there are 120 possible orders for matching target persons and protocols. Only one of these matchings is the true score, that is, the sum of all correct statements when all protocols are rated by the persons for whom they were intended. The other scores are control data. Hence, the significance of the true score can be evaluated based on the mean and standard deviation of the control values. This publication drew a lot of attention because it seemed that for the first time the problem of evaluating verbal material was solved. It turned out later, however, that the method inflates the results with a small number of sessions (Greville, 1949; Thouless, 1949). The reason is that with small amounts of data, the normal distribution, which the method uses for establishing the level of significance, is not an acceptable approximation of the true distribution. According to Greville (1949), the formula will usually be sufficiently accurate for data consisting of 10 or more sets of 5 records each.

The Pratt-Birge method was first applied in 1953 to evaluate an experiment carried out by an ASPR study group under Alan F. MacRobert (1954). Five sessions were held with different absent target persons who

were unknown to the psychic. The results were disappointing as the number of correct statements actually turned out to be below chance expectation.

Further British Research

D. J. West published the results of a two-year study with mediums (West, 1949, 1954). The design and execution was thorough: Target persons were not present during the sessions, and the person recording the statements was kept blind as to the target person's identity. Twelve mediums participated in 97 sessions. Two methods of evaluation were applied. In one, all target persons received all protocols and were requested to select the one that most applied to them. This evaluation yielded a nonsignificant negative result. As this method might be rather insensitive, West also applied an evaluation based on asking target persons to choose between the real statements and comparable controls. The sets were constructed in such a way that interdependency between statements could not bias the selection task. This method was applied to 15 sessions involving six mediums. The outcome was as disappointing as that of the first analysis: a nonsignificant negative deviation was again obtained. West noted how subjects differed in their ways of judging: Some agreed with nearly everything that the medium said, whereas others tended to deny most statements.

In 1951, an original variation of research with mediums was applied by V. M. Austin and was published by Tyrrell (1951). Austin presented a medium with an object and asked for precognitive statements. A target person was randomly selected from a pool of 30 participants, and this person was asked to carry the object for a year. At the end of the year, all participants judged the statements on applicability that had been made a year before. No objective evaluation was attempted, and the controls, when judging the statements, knew that the statements had not been intended for them. The results were difficult to interpret, but as Tyrrell noted, the use of precognitive statements is advantageous in making objective evaluations, and this should be exploited in subsequent research.

Research in Germany

In Germany, the problem of the objective evaluation of mediumistic material became the focus of attention primarily as regards the performances of the Dutch psychic Gerard Croiset. Croiset's precognitive chair tests, in which before a meeting he would describe the persons who would sit in certain chairs, lends itself well to research. In 1957, Bender published his first account of this kind of study, which he had been carrying out since 1953. He started with a series of exploratory studies, of which the last, the Pirmasens experiment, was described in the most detail. The experience gained with these studies was then applied in the next phase,

which was the development of quantitatively evaluated experiments with Croiset. Bender (1957, 1981) described a number of possible sources of error in such experiments. An important one was the *Zustimmungsbereitschaft,* or the tendency of target persons to agree with the statements of the psychic. Bender's method of evaluation was aimed at eliminating the influence of such sources of error.

The experimental procedure involved Croiset in precognitively describing one or more persons who would occupy randomly selected chairs at a future meeting. These statements were rated on applicability by all persons who attended the meeting before the target chair(s) were known. Based on the frequency of endorsements, each statement was assigned a weight for specificity. The more people who agreed with a statement, the less specific it was considered to be. A total score was then calculated for each respondent based on the correct answers multiplied by the weights for the statements. Based on the distribution of the scores thus obtained, it could be established, first, whether a person obtained a score that lies outside this distribution, thus suggesting that this might be the target person. Second, it could be decided whether that person occupied the chair Croiset had in mind. A useful aspect of this method, therefore, is that it also took into account possible displacement, that is, when Croiset described not the target person but someone else who attended. That displacement could occur had already been noted during the exploratory phase. This method only works, of course, as long as the medium or psychic cannot know beforehand which persons are going to attend.

The relevance of the latter statement had already appeared in the first quantitatively evaluated lottery-test of June, 1955. Lottery-tests are like chair-tests, but instead of people selecting their own chairs, they are randomly assigned numbers that designate their chairs or in some experiments, Croiset described in advance the specific number a subject would obtain. Before the numbers were assigned, Croiset described three participants. The first series of statements applied to a very significant degree ($z = 4.58$, $p \ll .01$) to a Dr. Neuhausler who, however, had obtained a wrong number in the lottery. It is not known to what extent Croiset could suspect that Neuhausler would attend. The second series was a clearly significant success ($z = 2.53$, $p = .01$), and the person to whom the statements best applied also drew the right number. The third series was an equally clear failure.

A second lottery experiment consisting of four trials was conducted a few months later that yielded equally mixed results. Two descriptions failed; for each of the other two descriptions, a participant was found with a score that significantly deviated from the distribution of scores for all participants ($p < .01$). However, both got wrong numbers. As a check, the statements of these two series were also judged by a comparable group of persons. This time no significant deviations were observed. Bender concluded that despite the first two failures, the experiment still yielded suggestive evidence for Croiset's precognitive ability.

Bender's method of evaluation still left room for improvement, a task taken up by the statistician Ulrich Timm, who developed a statistically valid method for the evaluation of psychic statements (Timm, 1965). Much care was given to the itemization of statements. For instance, the statement "The target person drives a red Volkswagen" would be split into "the person drives a car"; "the person drives a Volkswagen"; and "the person drives a red Volkswagen." Based on the weights derived from the responses from all participants, a *V*-value was derived for each statement. The total score for each participant is the sum of the *V*-values for all correct statements, and these total scores yield a normal distribution by which the significance of the score of the target person can be estimated.

Timm applied his statistical analysis to the data of new chair-tests with Croiset. In these chair-tests, (a) Croiset gave his statements without knowing which persons would be invited to attend; (b) invitations to persons were based on an objective random procedure; and (c) all participants rated the statements on applicability before the chairs were assigned. In the first experiment, Croiset predicted details about two persons who would attend. The results were mixed. One description fit the target person to a highly statistically significant degree (control distribution $M = 8.1$, $SD = 6.7$; the score the target person obtained was 55). The other prediction yielded a positive but nonsignificant result.

Apparently stimulated by this result, soon after a new series of precognition experiments with a slightly different procedure was carried out (Timm, 1966). An important variation was that in this experiment, all statements were first scored by a large group of persons via a questionnaire from which a smaller group was then randomly selected to attend the meeting. Four studies were conducted with 64, 57, 33, and 20 participants, respectively. Croiset knew beforehand the sex of the participant and that only students would be invited. However, the researchers soon discovered a new source of error. The participants' ratings of Croiset's statements differed according to whether the questions were asked via questionnaires or oral interviews. In oral interviews, more statements were accepted as applicable than when a questionnaire was used. Apparently with questionnaires, in cases of doubt respondents were inclined to reject statements, whereas in oral interviews they spent more effort on trying to remember incidents in their lives to which Croiset's statements might be related. If the oral interviews were used, 2 of the 4 experiments yielded statistically significant results. However, this outcome was clearly biased because the scores based on oral interviews with the target persons were evaluated against a distribution of scores from controls obtained by questionnaires.

The experience gained in these studies led to improvements in the procedure. For example, the questionnaires were formulated after first interviewing the target person to insure that the formulation of the statements represented the interpretation the target person gave to them. Three new studies were carried out. To avoid a possible displacement effect in the first study, only one person could be the target person. The results were so poor

that no further analysis was attempted. The second experiment involved two randomly selected target persons. After interviews with them, the questionnaires were constructed and submitted to 50 control subjects. However, it is unclear from the report whether these controls knew that they could not be the target person. Nevertheless, the results were very significant (combined $p < .0005$). The third study again involved only one possible target person. Although the statements seemed to apply to this target person, the result was not significant. Two of the 50 controls obtained a higher score. Hence in all, two of the seven experiments were clearly significant ($p < .01$), two were suggestive (significant when oral interviews were used, not significant when questionnaires were used), and three were not significant. Timm concluded that Croiset had demonstrated some paranormal ability, but that at best he seemed able to provide only a modest amount of factual information about target persons.

Further American Studies

One of the advantages of Bender's evaluation technique is that the specificity of statements is taken into account. It did not take long before attempts were made to improve the Pratt-Birge method along the same lines. Schmeidler (1958) applied weights in her study with the medium Caroline Chapman. Before the target persons judged the statements, Schmeidler classified them into 10 categories of specificity. Sixteen sessions were held in which the medium gave her impressions about absent target persons. Using weights and multiplication of weights for independent statements belonging to the same category, the Pratt-Birge method yielded a significant result ($p < .003$, uncorrected for multiple analyses). Applying the weights without multiplication or leaving out the weights reduced the results to a nonsignificant level. Schmeidler did not know at the time that the Pratt-Birge method inflates the result in the case of a small number of sessions. According to Schmeidler, the medium seemed most successful in those areas about which she was also normally most knowledgeable.

Roll (1962) offered a number of suggestions for quantitative methods for use in research with token-object tests, billet-reading tests, and ESP-projection tests. A few years later he applied these methods in studies with mediums. Together with Tart (Roll & Tart, 1965), he ran a series of exploratory tests, all variations on the theme that the psychics should be able to distinguish between pictures enclosed in envelopes with either strong or weak "strength of vibration." In the six experiments, no significant effects were observed.

Osis (1966) applied an innovative approach to the old problem of survival after death in his so-called "linkage" experiments with mediums, in which a chain of human links was interposed between the mediums and the widow of the deceased target person. The method was instituted in order to block possible ESP transmission. Using the mediumistic messages about

two deceased persons and cross-correspondences, his method involved paired comparisons, which allowed for the evaluation of the data by analysis of variance. However, apart from two marginally significant interaction effects, no indication of significant scoring was observed.

Schmeidler (1966) introduced an improved, quantitative approach to haunting cases. When members of a family reported that certain places in their home were haunted, Schmeidler asked them to mark on a floor plan the places the "ghost" favored and to score a personality checklist for traits that were characteristic and uncharacteristic of the ghost. Nine sensitives were then asked to tour the house. They marked copies of the floor plan for locations they considered haunted and scored the checklist on characteristics of the ghost based on their impressions. Two of the sensitives' locations were correct to a significant degree, and four other sensitives gave descriptions similar to those of the family.

Roll (1969, 1971) continued his research with psychics, but various tests with two mediums and 14 target persons yielded little success. Of interest is that he also investigated whether psychics could be helpful in identifying faces of unknown persons for possible application in police work. Applying the Burdick-Roll method (Burdick & Roll, 1971a, 1971b; Roll & Burdick, 1969), an improvement on the Pratt-Birge, the results were in all respects nonsignificant. He also continued his token-object tests, aimed among other things at testing his psi field theory (Roll, 1966a). Prior to the experiment, pairs of blank cards and metal chains with pendants were distributed to volunteers, who kept them for a specified period. The pairs were then separated, and the sensitive was requested to do a matching task. In none of the four experiments did the results deviate significantly from mean chance expectation. A third series of seven experiments (Roll, 1966b), with a similar design and again using cards, on the whole yielded positive scores, but they were not statistically significant. The only remarkable finding was that one subject consistently scored significantly on the cards of the same volunteer.

Tart (1969) reported a study with two female mediums in which the Pratt-Birge method was combined with "rarity-ratings" and yielded a nonsignificant outcome. No further details were given. Tart and Smith (1968) also ran two token-object studies with the Dutch psychic Peter Hurkos, who gave readings on envelopes containing hair from various cooperators. The results were evaluated by the Pratt-Birge method, again applying "rarity-ratings." However, it was found that judges were very inconsistent in assigning weights. No evidence for psi was found.

Stevenson (1967, 1968) and Mayne (1971) tried new methods of research or evaluation, but with little success, although Stevenson speculated that the high proportion of correct statements he obtained with a French medium could have been due to extrasensory perception. No statistical analysis was attempted.

Esser and LeShan (1969), having learned of Croiset's fame, attempted a chair-test, but the execution of their experiment allowed for nearly all the

sources of error that Bender had so painstakingly tried to avoid. No quantitative assessment appeared possible.

Stimulated by Schmeidler's work, Thelma Moss (Moss & Schmeidler, 1968) investigated a haunted house in Los Angeles. She followed Schmeidler's location chart and checklists for characteristics of the "ghost," but in addition to a group of psychics, a control group of nonpsychics was employed to check for the presence of stereotypical impressions. The members of the control group did not visit the house, but they filled in the forms as if they had seen a ghost in a house having such a floor plan. Three of the six sensitives showed a significant correspondence to the owners' descriptions, against only one of the control group of eight persons. The authors claimed that the results could not be explained by common stereotypes (assuming that psychics have stereotypes similar to nonpsychics). However, they do not know whether to attribute the results to observations of an actual deceased person or to ESP.

Roll, Morris, Damgaard, Klein and Roll (1972, 1973) and Damgaard (1972) continued their research with the medium, Sean Harribance. Readings were held in which pictures of target persons enclosed in opaque envelopes served as inductors. However, Roll discovered the same problem that Bender did: displacement effects. The design was aimed at reducing this source of error. Harribance provided readings on 20 target persons. When subjects were asked to select their own protocol, 6 out of 20 choices were correct, a weak significant result with an associated exact binomial p-value of .02. Applying the Pratt-Birge method yielded a nonsignificant outcome. Introducing weights for specificity (by using the same method Timm applied) again resulted in a weak significant effect ($p = .02$).

A series of precognitive dream studies with a young English psychic, Malcolm Bessent, were more successful. Bessent scored significantly in a precognition experiment run by Charles Honorton, in which he had to guess which of two colored lights would be lighted next by an electronically controlled random number generator. In two subsequent dream studies, one in 1969 and one in 1970, he scored even better. According to the authors, Bessent appeared to dream precognitively 14 out of 16 times (Ullman, Krippner, & Vaughan, 1989).

In the same period, Dean (1972) ran a study that was reminiscent of Austin's research (Tyrrell, 1951). It was aimed at reducing the notorious unreliability of judging by target persons. He asked 94 target persons to each provide three questions about the outcome of future events that were answerable by a yes or no. A medium provided the answers precognitively. The results were significant when for each correct answer a probability of .5 is assumed (which of course is not correct), but further details on the level of significance are not available. More impressive is the finding that no correlation appeared between the answers predicted by the subjects themselves (and they were in the best position to predict) and those of the medium, the medium being the better predictor.

In 1973, Eisenbud reported a chair-test replication with Croiset, who provided predictions for two target persons who would attend a demonstration evening a few weeks later in the U.S. Forty subjects took part from among whom the two target persons were selected by lot. However, although for both series of statements a person was found who could corroborate some of them, errors in the randomization procedure prevented knowledge of who the designated target persons actually were. The experiment was marred by many problems, among them that of translation, which was of sufficient seriousness that eventually a proper statistical evaluation could not be made. According to Eisenbud, a statistical evaluation was not required for the rich material provided by a brilliant psychic such as Croiset, an opinion in which he clearly differs with Bender (1957).

Schmeidler and Goldberg (1974) reported on a group experiment with one psychic and six nonpsychics, all of whom provided readings using inductors from four target persons. The only positive result was at the .05 level of significance for the reading of the psychic.

Another haunting case was reported by Maher and Schmeidler (1975) in which four psychics and eight skeptical nonpsychics attempted to identify the locations of appearance of the ghost and its characteristics. The results were not impressive. One psychic, to a weakly significant degree (p = .04), was able to identify spots of appearances but failed as regards characteristics. With another psychic, the opposite results were observed, with a p = .03 for mentioning correct characteristics of the appearance.

A similar study, applying the method that originated with Schmeidler, was reported by Maher and Hansen (1990, 1992). Four sensitives and four skeptics took part. Four of eight records collected from sensitives showed a weak or significant relationship to what witnesses had reported, but the pooled scores from the sensitives were not significantly different from those of the skeptics. The authors do not favor a specific interpretation for the results.

Haraldsson and Stevenson (1974) conducted an experiment with the Icelandic medium Hafstein Björnsson. In a random order, 10 sitters individually were brought to an experimental room where the medium was visually separated from them by an opaque curtain. The sitters were acoustically isolated from the medium. The medium's impressions of deceased persons, usually relatives or friends, which he claimed to "see," were recorded for each sitter. Later, each sitter was given all 10 reports and instructed to rank them according to how clearly he or she could identify the persons described. Four sitters identified the reading given for them, which yielded a significant p = .01. However, 11 further experiments with the same medium reported four years later (Haraldsson, Pratt, & Kristjansson, 1978) failed to confirm the positive results of the initial experiment.

In unsolved cases, the police frequently deal with purportedly paranormally acquired information provided by psychics. Reiser, Ludwig, Saxe, and Wagner (1979) and Reiser and Klyver (1982) carried out a study in

which psychic impressions were obtained about four crimes, two solved and two unsolved. In the second study, impressions were obtained from two control groups, one composed of detectives and the other of students. Only verifiable statements were included in the analysis. Although the methodology of the analysis was not optimal, the data do not indicate that psychics are able to provide useful information in criminal investigations. No differences were observed between the group of psychics and the control groups.

A last chair-test with Croiset was carried out by W. Kugel (1979) of Berlin. Unfortunately, the number of participants was much smaller than the investigator had anticipated: 53 instead of the expected 240. Croiset made advanced predictions about two target persons. All subjects rated these statements on applicability to themselves before the target persons were identified. Kugel did not attempt to apply Timm's method of evaluation, but he decided in advance that the scores of the target persons would have to fall in the range of the 5% highest scores to be considered significant. This was not the case, nor was a displacement effect observed.

Bender (1981) published a review of the exploratory Pirmasens study (see Murphy, 1957) and added new data based on later interviews with target persons. Hoebens (1984) reacted with a highly critical—and in my opinion, in many respects unjustified—attack on Bender and the Pirmasens study. This motivated Timm (1984) to run a control study to provide a quantitative evaluation of the Pirmasens data. Seventy-two and 89 control subjects, respectively, rated the statements of the two predictions of that study on applicability. Based on the distributions obtained, it appeared that in both cases Croiset had described the target persons with marginally significant accuracy ($p = .02$ and .04, respectively). Timm stated that this result had to be taken with reserve because the controls knew the statements were not intended for them.

Recent Dutch Studies

Bosga and Perizonius (1987) reported a study in which four psychics gave impressions using bones as inductors. The oldest dated back to 10,000 BC; the most recent concerned a skull with a shot wound. Of the 139 statements, 44 were rated by the authors as verifiable. Based on a subjective estimation of the probabilities of these statements by the authors, 20 were rejected because the correctness of the statements might have been due to chance. Of the remaining 24 statements, seven proved to be correct and hence, according to the authors, might be considered to be paranormal in origin.

In Holland, largely due to the work of W. H. C. Tenhaeff, there still exists much interest in engaging psychics to trace missing persons or for use in criminal investigations. Neu (1985) analyzed 112 paranormal "tips" that were received from psychics after Heineken (the director of the famous brewery) and his driver were abducted in 1983. The police solved

the case in 18 days. To his great dismay, Neu found that hardly any of the paranormal impressions corresponded to the real situation. Only three of the tips contained elements that related to aspects of the case. However, according to the leader of the investigating team, even if these three tips had been acted upon, they could not possibly have contributed to the solution of the case. Neu concluded that the contributions of psychics to criminal investigation need supervision by experienced parapsychologists in order to be of any use.

An occasion to prove this view came a few years later with the abduction of G. J. Heijn, a prominent captain of industry, in 1989. This case lasted much longer, about seven months, before it was solved. Because of its long duration and the fact that Heijn's business offered a substantial reward to the person providing information leading to Heijn's whereabouts, over 1,500 paranormal tips were received. The researchers (Gerding, Millar, Molewijk, Neu, & Voois, 1989) attempted to correlate these tips, looking for patterns that might suggest Heijn's location. This largely failed, partly due to continual problems involved in obtaining the needed computer facilities and partly because the tips were so varied in nature that no pattern could be found. Psychics were also actively engaged in the case. Thirteen psychics gave their impressions about the fate and location of the victim, but they differed strongly among each other, and they did not provide any worthwhile clues. Psychics also handled objects that had belonged to Heijn and then tried to reconstruct the route the abductors had followed. These attempts also failed. The best result was obtained from one psychic who drew a small hill with a gallows on it. Later it was discovered that 12 km from the location where eventually Heijn was found a "Gallows Hill" existed. However, the name is not uncommon in Holland, and the psychic did not point out the location of this "Gallows Hill." After the case was solved, it turned out that the crime was committed by only one person, unforeseen by anyone, who had killed Heijn the very day of the abduction. A post hoc analysis revealed that not even one tip or psychic attempt came close to what happened.

By far the most extensive experimental study with psychics was carried out by Boerenkamp (1988a, 1988b). This study differed in important respects from the usual studies with mediums and psychics. In the standard condition, psychics gave their readings under the same circumstances that they used when giving readings for clients, including feedback from the sitters. In nearly all other studies, feedback was avoided in order not to compromise a quantitative evaluation. In addition, Boerenkamp systematically varied important experimental variables, such as degree of feedback, presence of the target person, importance of the events about which the psychic was to provide information, and the nature of the inductor. All sessions were held in the homes of the psychics. The aims of the study were: (a) to determine if psychics give a sufficient number of correct statements to justify the conclusion that experimentation with them is to be preferred over experimentation with unselected subjects; (b) to provide a

description of what psychics do and what readings they give when they are consulted by a client; (c) to study the effect of manipulation of variables on the psychic's achievements; and (d) to compare the achievements of psychics with those of nonpsychics who give impressions under comparable conditions. The second aim was included because nearly all quantitative studies of mediums and psychics have concentrated on ESP phenomena. By including the third aim, it was hoped that conditions could be found that would increase the ESP effects.

Normally in studies with psychics, the aim is to demonstrate that psychics are able to provide more information on target persons than can be expected by chance. However, in most studies the experience of psychics in dealing with persons not known to them cannot be ruled out as a contributing factor when positive results are obtained. In my view, therefore, the proper question, and the only one of practical use, is not whether psychics are able to do better than chance, but whether psychics are able to do better than nonpsychics of comparable experience in dealing with target persons.

The design of the study, partly because of the inclusion of feedback conditions and because of the enormous amount of data, called for a different evaluative technique than those used in previous studies. Boerenkamp called it the *window model,* because all statements are first rated by independent judges on spontaneity and specificity, and only statements that were judged to be sufficiently spontaneous and specific were considered as having potential paranormal value and so were included in the final evaluation. Spontaneity involved statements whose content is not related to previous statements or feedback. Hence, even nonspecific statements might obtain a degree of spontaneity sufficiently high to pass the threshold for inclusion in the final analysis. Basically, the method excluded from the final analysis all the "rather likely" statements, in view of the information available to the psychic at the moment the statement was made, thereby reducing the database to manageable proportions. The choice of cut-off criteria for specificity and spontaneity was rather arbitrary. In Boerenkamp's study, a 10% level was maintained; that is, only 10% of the total statements were used for the evaluation. Because this 10% consisted of the most spontaneous and specific statements, it was hoped that they would provide a "window through which evidence for ESP could be seen" (Boerenkamp & Schouten, 1983).

Twelve psychics gave 13 readings, each under different conditions. In addition, two groups of 12 nonpsychics participated. One group was comparable to the psychics in terms of their experience in handling other people's problems, and the other was comparable in terms of sex, age, and educational background. The study involved over 200 sessions and over 10,000 statements. From an analysis of the content of the readings, it appeared that each psychic had a personal preference for the length of a reading that remained stable over the 5-year period of data gathering. But the length of the reading was strongly influenced by the conditions. It

varied from an average of 144 statements, in cases where the target person was present and providing full feedback, to an average of 17 statements where the statements were made without any social context, including feedback (i.e., when the psychic handled the inductor at home at a time he or she felt most optimal). Thus, in addition to individual differences, paranormal ability, as expressed by the number of statements, is strongly influenced by external conditions. As a rule, one can state that the more sensorial information that is available, the higher is the number of statements. Nonpsychics provide significantly fewer statements.

The number of potential paranormal statements that turned out to be correct was approximately 1%, a value that could be expected by chance. Nonpsychics achieved comparable results. This finding is in agreement with the results of Reiser et al. (1979, 1982). Boerenkamp also observed no difference between the achievements of psychics and controls. Especially in the case of rather specific questions and statements, the same phenomenon was observed as later in the Heijn case: Nearly all psychics strongly disagreed. None of the conditions investigated had a noticeable positive effect on the performance, and no differences in paranormal abilities were observed between psychics. However, a difference between psychics was observed in number of potentially paranormal statements. Some psychics appeared to take more risks by making more specific statements than others. In addition, a content analysis on eight different criteria (topics; persons; statement related to past, present, future; statements about favorable, neutral, and unfavorable states of affairs; statements in the form of advice; a number of statements preceded by silence; nonrhetorical and rhetorical statements; and numbers of statements preceding a feedback action) gave no indications that with any of them psychics did better in providing correct information.

Most statements were about the personality and psychological conditions of the target person, followed by work, health, and relationships with relatives. Again, external conditions, such as degree of feedback or seriousness of event, had no significant influence on these distributions. Nonpsychics clearly differed in that they made fewer statements about psychological conditions and health problems. Another difference is that nonpsychics restricted their statements to the target person, whereas psychics tended also to make statements about persons related to the target person. It was also found that when target persons were present during the readings, they gave by far the highest proportion of feedback—substantially more than sitters who provided feedback on absent target persons. This might be the reason that psychics prefer to give readings in the presence of the target person. When psychics made a statement and were confronted with a denial from either the target person or the sitter, the most common reaction, observed in 60% of these cases, was to give their statement a new interpretation. The simple acceptance of the denial occurred in only about 10% of these cases.

CONCLUSIONS

It appears that the development of adequate quantitative evaluation methods for mediumistic statements, which we now have, has been not only a matter of developing the proper statistical methods but also a gradual discovery of many potential sources of error. This means that earlier research in which the authors claimed to have achieved positive results might present too optimistic a picture, because in these studies not all sources of error were eliminated. On the other hand, although contemporary evaluative methods (such as Burdick-Roll and Pratt-Birge) are statistically sound, they are not applicable to the situation that psychics most commonly work in: reading with feedback. In that respect, in most of the investigations reported the psychics had to work under adverse conditions. However, especially in view of the recent Dutch research, it is doubtful whether this would have made much difference.

It can be concluded that most research on psychics yielded nonsignificant results. In the most optimistic scenario, even counting studies like those of Thomas as successful, less than half of the studies reported positive results. The real success rate, that is, the rate if the significances were corrected for the various sources of error in the studies, is probably closer to one out of three studies reporting some significance. In addition, where the results turned out to be significant, they were often marginal and not impressive. Even with a star subject such as Croiset, most experiments failed, and the successful ones rarely exceeded the .01 significance level. In comparison with laboratory research with mainly unselected subjects, for instance the Ganzfeld studies, studies with psychics clearly are not more successful. Although, in principle, anyone may call him- or herself a psychic, with few exceptions the survey material is based on work with well-known psychics. The experimenters, at least, had sufficient confidence in the psychics' abilities to invest considerable time and effort.

This result is clearly at variance with the popular image of the abilities of psychics. If that image is not based on demonstrable ESP ability, as the data suggest, then the question remains as to why so many people are impressed by what psychics do. I believe that an important reason for this is that this image is mainly based on a few spectacular cases, often rather selectively and incompletely reported by the media. The sometimes year-long studies of psychics and their failures are rarely mentioned. The popular image of the psychic, at least in the Netherlands, is often based on a few highly publicized cases concerning a small number of "stars," together with the endless and often distorted repetition of these few cases in the popular literature. This image is probably reinforced by what common people experience when consulting a psychic.

In the following section I describe what I think takes place when a psychic is consulted for advice and gives a reading to a client. Various psychological processes might explain why it is likely that even without

any paranormal impressions both the psychic and the sitter will attribute a paranormal character to the psychic's statements. This interpretation is based on the data of the studies discussed above and on my own experience in the many sittings I have had with psychics.

An Interpretation of the Way Psychics Operate

A reading is basically a verbal interaction between a psychic, who claims paranormal abilities, and a client, who consults the psychic, usually for advice on a specific problem. Typically, after some initial conversation, the psychic signals, either behaviorally or verbally, that he or she has entered a state in which paranormal impressions can be obtained. It is then assumed by client and psychic that what follows is paranormally obtained information. An interaction develops in which frequently the paranormal character, that is, the truth of the statements, is established. It is striking that most statements of the psychic concern verifiable matters, that is, knowledge the client already possesses. In between, the psychic gives impressions or advice on the unverifiable matters the client came for, sometimes mixed with opinions on subjects of a more general nature, such as ethical and religious matters, how to live, and so forth.

Although the psychic acts as if a clear switch is made from "normal" to "paranormal," I propose that in both states mainly normal psychological processes are involved in the interaction. It might well be that occasionally paranormal elements emerge. However, the difference between the two states, or the two roles of the psychic, seems to me of more importance for the nature of the psychological processes involved. As long as the psychic acts as if in the "normal" state, the interaction is like any other conversation between people in similar circumstances. But from the moment the statements of the psychic are labelled "paranormal," the nature of the interaction changes. The psychic has to establish authority by providing information which, according to the client, he or she cannot normally know. In order to do this, information has to be verifiable. The psychic therefore rarely restricts him- or herself to just telling the client what he or she came for. The bulk of the interaction consists of statements and feedback about matters familiar to the client. In addition, there is another important reason for spending time on matters the client already knows: It serves as an important source for generating feedback, that is, information from the client. This enables the psychic to form an understanding of the background of the problem. Therefore, most statements of psychics are, as Boerenkamp (1988b) called them, of a rhetorical or open-ended nature. They are not just statements, but rather they are formulated in such a way that they stimulate feedback.

The role of the client in this interaction is often underestimated. Normally the focus is on the psychic, whereas the client is assumed to play a passive role. But in the "paranormal" phase, the client becomes active, too. Most clients turn to a psychic because they have a problem that they

can't solve by other means. In addition, clients seek advice, and even if they still have some reservations about the paranormal abilities of the psychic, which I think is often the case, they will try to prevent a failure. Hence, once the psychic is in the paranormal phase, the client takes an active role in trying to confirm the paranormal character of the verifiable statements by finding facts or interpretations with which they fit. One might say: A psychic is as able as the client allows! Both partners cooperate to make the statements true. Often the psychic helps the client. If the client can't find a match, the psychic may offer a different interpretation for the statement. Then the client starts the search process again.

I am convinced that in most cases neither the psychic nor the client is aware of their respective contributions and the psychological processes involved. In addition, a number of other psychological processes, well known from the professional psychological literature, contribute to the apparent success of a sitting. These are:

1. In general, the psychic controls the encounter to a much greater extent than the client is aware of.
2. Many topics for readings concern basic needs or fears that almost all people experience. General statements in these areas that create belief or assurance will often be considered as "correct" and as very applicable to the client's situation.
3. The appearance of the client provides the psychic with information about the person.
4. Clients might not be aware of their nonverbal reactions to statements on topics that are highly emotional for them. The psychic might not be aware that he or she processes such nonverbal feedback.
5. Clients have a common tendency to avoid disagreements and dissonance by trying to confirm statements (*Zustimmungsbereitschaft*).
6. People tend to attach more value and to remember better correct statements rather than incorrect ones, especially in cases where a search process finally yields a satisfactory interpretation. The search process itself and the initial false interpretations are quickly forgotten.
7. People do poorly at estimating probabilities, and they tend to underestimate the probability of correct statements.
8. Clients have a tendency to attribute a paranormal character to all statements made in the "paranormal" phase of the sitting, including simple ones to which in normal conversation they would attach no importance.
9. Most statements have a variety of interpretations. This not only strongly increases the probability of being correct, but it also creates a situation in which a client may affirm a statement based on an interpretation that is different from that which the psychic intended; that is, the client "reads in" the correct answer.
10. Clients may strongly underestimate the experience of psychics. Because clients accept the "paranormal" state of the psychic, they seem to assume that in this state normal psychological processes, such as the effect of experience and rational inference, are excluded. This is certainly not the

case because one can clearly see the effect of the feedback given by the client on subsequent statements of the psychic, which demonstrates that normal information processing functions are operating in the "paranormal" state.

11. To a client, his or her own problems and situation are unique, but that does not apply to the psychic. Most questions and problems that clients present to them are rather common. Thus, a standard repertoire of possible answers often can answer any given client's need.

12. An interaction is an active process from which only those elements will be remembered that are of special importance. For the client, the impression made by the psychic is foremost. Hence, clients are inclined to underestimate and to forget their own contribution to the interaction, including the comments and feedback they provided.

In summary, the psychic/client interaction is an active process that leaves little time for reflection. Most clients and psychics will not even be aware of the possible effect of the 12 normal psychological processes listed above. It is therefore not surprising that often both parties are honestly impressed by the results of a reading. This reinforces the belief of the psychic in his or her own abilities and convinces the client that a psychic can do things that a normal person cannot. For both it reduces the unavoidable uncertainty that is inherent in the psychic-client relationship. That psychics occasionally have real paranormal impressions certainly cannot be excluded, but from the available data we can conclude that if it is present, it plays a very minor role. Moreover, a paranormal faculty does not have to be assumed to explain the success of the interaction between psychic and client.

It is of interest to compare the readings of psychics with the spontaneous paranormal experiences of normal people. From my experience with both types of data, I have the impression that spontaneous experiences are much more impressive than the readings of psychics. Of course, in the case of spontaneous experiences, the less impressive and trivial occurrences are likely to be less often reported. Still, this seems to me insufficient to explain the difference. Boerenkamp's (1988b) research suggests that psychics behave stereotypically, as was earlier noted by Saltmarsh (1929).

I believe that acting as a psychic actually has a negative effect on whatever paranormal experiences a psychic may have had originally. My impression is that most psychics have had a difficult youth and a rather unsuccessful adult life, which provides them with an above-average insight into the psychological problems and circumstances of other people. In this respect, they may be real "sensitives." In addition, they have had quite a number of "strange" experiences, not necessarily all of a psi type, from which they and others conclude that they have paranormal ability. However, these experiences are basically spontaneous and unpredictable. From experimental research, such as the Ganzfeld experiments, we know that even under the most optimal conditions the amount of ESP information is limited. A person becomes a psychic most often not because of a conscious

decision but rather, because he or she is gradually stimulated by social support from the environment. However, the more professional psychics become, the less are they able to wait for spontaneous impressions. Clients want answers immediately, so psychics are forced to rely on experience and sensorially obtained information to provide answers. This becomes routine and may well eventuate in the loss of or a blocking of whatever paranormal sensitivity they may have had initially. In other words, professional psychics could be even less paranormally sensitive than nonpsychics. Malcolm Bessent (see Ullman, Krippner, & Vaughan, 1989) seems to be an exception, as he scored so well in the Maimonides dream studies. But in those days he was a young psychic and probably had had little experience as a professional.

If this hypothesis is correct, it does not imply that psychics are dishonest or worse, or that it is senseless to consult a psychic. Psychics and clients must both be honestly impressed on occasion by the achievements that confirm the psychic's belief in his or her abilities. In addition, one should not underestimate the experience psychics might gain in their profession. Undoubtedly, some psychics who specialize, for instance, in locating missing persons, may have a much larger experience with such cases than the average police officer. Whatever the source of their statements, their experience and expertise in certain areas might make it worthwhile to consult them for advice, even if they were to make no genuine paranormal statements in a given session. But the criterion to use in selecting a psychic seems to me to be the expertise and experience of the person in regard to the type of problem for which the client is consulting, rather than an estimate of his or her supposed paranormal sensitivity, as indicated, perhaps, by some standard ESP test.

References

BENDER, H. (1957). Präkognition im Qualitativen Experiment: Zür Methodik der "Platzexperimente" mit dem Sensitiven Gerard Croiset. *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie, 1,* 5–36.

BENDER, H. (1981). Pirmasens 1953—Retrospektive auf ein Platzexperiment mit Gerard Croiset. *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie, 23,* 219–230.

BOERENKAMP, H. G. (1988a). *Helderziendheid Bekeken.* Haarlem, Holland: De Toorts.

BOERENKAMP, H. G. (1988b). *A Study of Paranormal Impressions of Psychics.* Doctoral dissertation, University of Utrecht. (Available in book form from the author.)

BOERENKAMP, H. G., & SCHOUTEN, S. A. (1983). Estimating the potential paranormal value of verbal statements. *Journal of Parapsychology, 47,* 121–130.

BOSGA, D., & PERIZONIUS, R. (1987). Extrasensory perception in an-

thropo-osteology: An exploratory experiment. *Bones: Treasuries of Human Experience in Time and Space, 1,* 73–93.

BURDICK, D. S., & ROLL, W. G. (1971a). Differential weighting of ESP responses. *Journal of the American Society for Psychical Research, 65,* 173–184.

BURDICK, D. S., & ROLL, W. G. (1971b). Differential weighting of verbal and other ESP responses with different probabilities [Summary]. *Proceedings of the Parapsychological Association 1969,* No. 6, 26–27.

DAMGAARD, J. (1972). Content analysis of free verbal response material [Summary]. *Proceedings of the Parapsychological Association 1970,* No. 7, 18–19.

DEAN, D. (1972). A precognitive method for testing sensitives [Summary]. In W. G. Roll, R. L. Morris, & J. D. Morris (Eds.), *Research in Parapsychology 1972* (pp. 97–99). Metuchen, NJ: Scarecrow Press.

EISENBUD, J. (1973). A transatlantic experiment in precognition with Gerard Croiset. *Journal of the American Society for Psychical Research, 67,* 1–25.

ESSER, A. H., & LESHAN, L. (1969). A transatlantic "chair test." *Journal of the Society for Psychical Research, 45,* 167–170.

FINDLATER, J. W. (1932). The statistical method of assessing the evidential value of mediumistic communications [Letter to the editor]. *Journal of the Society for Psychical Research, 27,* 275–276.

GERDING, J. L. F., MILLAR, B., MOLEWIJK, G. C., NEU, J. G., & VOOIS, W. (1989). Analyse van de paranormale tips in de ontvoeringszaak G. J. Heijn. *Tijdschrift voor Parapsychologie, 57,* 3–71.

GREVILLE, T. N. E. (1949). On the number of sets required for testing the significance of verbal material [Letter to the editor]. *Journal of Parapsychology, 13,* 137–138.

HARALDSSON, E., PRATT, J. G., & KRISTJANSSON, M. (1978). Further experiments with the Icelandic medium, Hafsteinn Björnsson. *Journal of the American Society for Psychical Research, 72,* 339–347.

HARALDSSON, E., & STEVENSON, I. (1974). An experiment with the Icelandic medium Hafsteinn Björnsson. *Journal of the American Society for Psychical Research, 68,* 192–202.

HERBERT, C. V. C. (1937a). An experiment with Mrs. Garrett. *Journal of the Society for Psychical Research, 30,* 99–101.

HERBERT, C. V. C. (1937b). Report on a series of sittings with Miss Frances Campbell. *Journal of the Society for Psychical Research, 30,* 2–16.

HETTINGER, J. (1940). *The Ultra-Perceptive Faculty.* London: Rider.

HETTINGER, J. (1941). *Exploring the Ultra-Perceptive Faculty.* London: Rider.

HETTINGER, J. (1947). Psychometric telepathy across the Atlantic. *Journal of the American Society for Psychical Research, 41,* 94–122.

HOEBENS, P. H. (1984). Abschied von Pirmasens. *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie, 26,* 4–28.

HYSLOP, J. H. (1919). Chance coincidence and guessing in a mediumistic experiment. *Proceedings of the American Society for Psychical Research, 13,* 5–88.

KUGEL, W. (1979). Ein Platzexperiment mit Gerard Croiset. *Arbeitsbericht 7 Forschungsgebiet Parapsychologie 1979.* Berlin, Germany: Technische Universitt Berlin.

MACROBERT, A. F. (1954). Proxy sittings: A report of the study group series with Arthur Ford. *Journal of the American Society for Psychical Research, 48,* 71–73.

MAHER, M. C., & HANSEN, G. P. (1990). Quantitative investigation of a reported haunting using several detection techniques. *Proceedings of Presented Papers: The Parapsychological Association 33rd Annual Convention,* 151–167.

MAHER, M. C., & HANSEN, G. P. (1992). Quantitative investigation of a reported haunting using several detection techniques. *Journal of the American Society for Psychical Research, 86,* 347–374.

MAHER, M. C., & SCHMEIDLER, G. R. (1975). Quantitative investigation of a recurrent apparition. *Journal of the American Society for Psychical Research, 69,* 341–352.

MAYNE, A. (1971). Some approaches to the quantitative evaluation of qualitative experiments in parapsychology [Summary]. *Proceedings of the Parapsychological Association 1968,* No. 5, 46–48.

MOSS, T., & SCHMEIDLER, G. R. (1968). Quantitative investigation of a "haunted house" with sensitives and a control group. *Journal of the American Society for Psychical Research, 62,* 399–410.

MURPHY, G. (1957). Triumphs and defeats in the study of mediumship. *Journal of the American Society for Psychical Research, 51,* 125–135.

NEU, J. G. (1985). De waarde(loosheid) van spontane bijdragen van paragnosten: Een analyse van paranormale tips in de Heineken-affaire. *Tijdschrift voor Parapsychologie, 53,* 2–23.

OSIS, K. (1966). Linkage experiments with mediums. *Journal of the American Society for Psychical Research, 60,* 91–124.

PARSONS, D. (1949). On the need for caution in assessing mediumistic material. *Proceedings of the Society for Psychical Research, 48,* 344–352.

PRATT, J. G. (1936, March). Towards a method of evaluating mediumistic material. *Bulletin of the Boston Society for Psychic Research,* No. 23, 5–53.

PRATT, J. G. (1960). Methods of evaluating verbal material. *Journal of Parapsychology, 24,* 94–109.

PRATT, J. G. (1969). *On the Evaluation of Verbal Material in Parapsychology* (Parapsychological Monographs No. 10). New York: Parapsychology Foundation.

PRATT, J. G., & BIRGE, W. R. (1948). Appraising verbal test material in parapsychology. *Journal of Parapsychology, 12,* 236–256.

REISER, M., & KLYVER, N. (1982). A comparison of psychics, detectives and students in the investigation of major crimes. In M. Reiser (Ed.),

Police Psychology: Collected Papers. Los Angeles: LEHI Publishing Company.

REISER, M., LUDWIG, L., SAXE, S., & WAGNER, C. (1979). An evaluation of the use of psychics in the investigation of major crimes. *Journal of Police Science and Administration, 7,* 1, 18–25.

RHINE, J. B. (1934a). *Extra-Sensory Perception.* Boston, MA: Boston Society for Psychic Research.

RHINE, J. B. (1934b). Telepathy and clairvoyance in the normal and trance states of a "medium." *Character and Personality, 3,* 91–111.

ROLL, W. G. (1962). Designs for tests with free response material. *Journal of the American Society for Psychical Research, 56,* 184–195.

ROLL, W. G. (1966a). Further token object tests with a "sensitive." *Journal of the American Society for Psychical Research, 60,* 270–280.

ROLL, W. G. (1966b). Token object matching tests: A third series. *Journal of the American Society for Psychical Research, 60,* 363–379.

ROLL, W. G. (1967). ESP tests with two mediums [Summary]. *Proceedings of the Parapsychological Association 1966,* No. 3, 7–8.

ROLL, W. G. (1969). Experiments with the medium, Douglas Johnson [Summary]. *Proceedings of the Parapsychological Association 1967,* No. 4, 15–16.

ROLL, W. G. (1971). Free verbal response and identi-kit tests with a medium. *Journal of the American Society for Psychical Research, 65,* 185–191.

ROLL, W. G., & BURDICK, D. S. (1969). Statistical models for the assessment of verbal and other ESP responses. *Journal of the American Society for Psychical Research, 63,* 287–302.

ROLL, W. G., MORRIS, R. L., DAMGAARD, J. A., KLEIN, J., & ROLL, M. (1972). Free verbal response tests with Lalsingh Harribance [Summary]. *Proceedings of the Parapsychological Association 1970,* No. 7, 21–23.

ROLL, W. G., MORRIS, R. L., DAMGAARD, J. A., KLEIN, J., & ROLL, M. (1973). Free verbal response experiments with Lalsingh Harribance. *Journal of the American Society for Psychical Research, 67,* 197–207.

ROLL, W. G., & TART, C. T. (1965). Exploratory token object tests with a "sensitive." *Journal of the American Society for Psychical Research, 59,* 226–236.

SALTMARSH, H. F. (1929). Report on the investigation of some sittings with Mrs. Warren Elliott. *Proceedings of the Society for Psychical Research, 39,* 47–184.

SALTMARSH, H. F., & SOAL, S. G. (1930). A method of estimating the supernormal content of mediumistic communications. *Proceedings of the Society for Psychical Research, 39,* 266–271.

SCHMEIDLER, G. R. (1958). Analysis and evaluation of proxy sessions with Mrs. Caroline Chapman. *Journal of Parapsychology, 22,* 137–155.

SCHMEIDLER, G. R. (1966). Quantitative investigation of a "haunted

house." *Journal of the American Society for Psychical Research, 60,* 137–149.

SCHMEIDLER, G. R., & GOLDBERG, J. (1974). Evidence for selective telepathy in group psychometry [Summary]. In W. G. Roll, R. L. Morris, & J. D. Morris (Eds.), *Research in Parapsychology 1973* (pp. 103–106). Metuchen, NJ: Scarecrow Press.

SCOTT, C. (1949). Experimental object-reading: A critical review of the work of Dr. J. Hettinger. *Proceedings of the Society for Psychical Research, 49,* 16–50.

SCOTT, C. (1972). On the evaluation of verbal material in parapsychology: A discussion of Dr. Pratt's monograph. *Journal of the Society for Psychical Research, 46,* 79–90.

STEVENSON, I. (1967). A new method for evaluating verbal material in mediumistic communication [Summary]. *Proceedings of the Parapsychological Association 1967,* No. 4, 14–15.

STEVENSON, I. (1968). The analysis of mediumistic session by a new method. *Journal of the American Society for Psychical Research, 62,* 334–355.

TART, C. T. (1969). Token object studies with two Californian psychics [Summary]. *Proceedings of the Parapsychological Association 1967,* No. 4, 16.

TART, C. T., & SMITH, J. (1968). Two token object studies with Peter Hurkos. *Journal of the American Society for Psychical Research, 62,* 143–157.

THOMAS, C. D. (1939). A proxy experiment of significant success. *Proceedings of the Society for Psychical Research, 45,* 257–306.

THOMAS, J. F. (1937). *Beyond Normal Cognition: An Evaluation and Methodological Study of the Mental Content of Certain Trance Phenomena.* Boston Society for Psychic Research.

THOULESS, R. H. (1949). Review of *Journal of Parapsychology,* V.12, N.4. *Journal of the Society for Psychical Research, 35,* 48–49.

TIMM, U. (1965). Die Statistische Analyse Qualitativer Paragnostischer Experimente. *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie, 8,* 78–122.

TIMM, U. (1966). Neue Experimente mit dem Sensitiven Gerard Croiset. *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie, 9,* 30–59.

TIMM, U. (1984). Eine Statistische Kontrolluntersuchung zum Pirmasenser Platzexperiment mit G. Croiset (1953). *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie, 26,* 29–51.

TYRRELL, G. N. M. (1951). An experiment in precognition. *Journal of the Society for Psychical Research, 36,* 366–368.

ULLMAN, M., KRIPPNER, S., & VAUGHAN, A. (1989). *Dream Telepathy: Experiments in Nocturnal ESP* (2nd ed.). Jefferson, NC: McFarland.

WEST, D. J. (1949). Some proxy sittings: A preliminary attempt at objective assessment. *Journal of the Society for Psychical Research, 35,* 96–101.

WEST, D. J. (1954). Experimental parapsychology in Britain: A survey of published work, 1948–53. *Journal of the Society for Psychical Research, 37,* 323–347.

University of Utrecht
Trans II, K. 1704
Heidelberglaan 2
3584 CS Utrecht
The Netherlands

Department of Psychology
University of Edinburgh
7 George Square
Edinburgh EH8 9JZ
Scotland UK

Part III
Extrasensory Perception

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[13]

EXPERIMENTS BEARING ON THE PRECOGNITION HYPOTHESIS:

I. PRE-SHUFFLING CARD CALLING

J. B. RHINE

ABSTRACT: This is the first of a series of reports by this author and by others, on researches into the question whether time is a limiting condition in the functioning of ESP. Certain ESP procedures were used to test whether a card order could be called in advance of shuffling and the calls then checked against the pack. After many preliminary tests (summarized herein), these "precognition" tests were administered to 49 subjects under especially guarded conditions in which subjects were not allowed to handle the cards or check results. In the course of the total 113,075 trials a deviation was given such as would not be expected by chance more than once in about 400,000 such lengthy series. Some hypotheses alternative to that of precognition are considered and a position of suspended judgment reached in which the precognition hypothesis is favored but not regarded as proved.

THE HYPOTHESIS that there is an extra-sensory mode of perception has by now passed a number of barriers to its establishment with a fair degree of success. These barriers have consisted mainly of three counter-hypotheses which should be considered as possible explanations of the results of the experiments made to investigate the question of extra-sensory perception. The first was that the experimental results were due to chance; the second, that they were due to sensory cues obtained from the test cards; and third that they were due to peculiar weaknesses in the research personnel. It has now been roundly recognized that the first of these, the hypothesis of chance[1] is quite inadequate to explain the results. Likewise the evidence obtained from the more advanced experimental conditions is showing the second alternative to be equally inapplicable as an explanation, for under these guarded conditions sensory cues from the test cards would be an impossibility. Meanwhile increasing repetition of the work on extra-sensory perception in other college laboratories is fast reducing concern over the third major question —that of the research personnel, by adding to the already considerable

[1] See statement by Professor Burton H. Camp, p. 305, December, 1937, issue of the JOURNAL OF PARAPSYCHOLOGY.

number of studies confirming the earlier work which was done in the Duke laboratory.

As the degree of general assurance as to the occurrence of this type of perception grows, it becomes increasingly possible to advance the inquiry further toward the discovery of the nature of extra-sensory perception or ESP. Toward this goal, those most familiar with the ESP research have long been impatient to advance. As a matter of fact, they have not been idle in this regard, and some important steps have already been taken.

In approaching the question of the nature of the ESP process or phenomenon, one may stress, first, its mental relations, and try to place it in the schema of what is already known of the mind. Or, second, its bodily relations, if there be such, may be sought after in terms of possible organic or physiological associations (or their absence). Finally, one may search after the more external, perhaps more general limiting and determining conditions of ESP such as its place (or lack of place) in the spatial and temporal schemata of events by which orientation in the physical world is achieved.

Beyond likelihood of question, every one of these fields of relations is of great importance. But of first consideration, if not of ultimately prime significance is the third problem, that of the relation of ESP to time and space, for the answer to this must prove a guide to the most fundamental distinctions which can be made in classification.

No sooner was the evidence from the ESP experiments past the criteria of statistical significance with sensory cues securely excluded than was distance between subject and stimulus introduced as a test condition. The consequences of this advance exploration were, as readers of this JOURNAL already know, sharply at variance with the general laws of mechanical nature as far as they are generally understood. As I have shown in a summary of the experiments in which distance was a condition of the tests[2] there is no effect of space yet discovered in the results of these tests, and it would appear from the data that no appreciable effect is likely to be found.

Now if the ESP process is not detectably affected by spatial conditions we may rightly inquire how it could then be detectably limited by time. For time measurements are made in reference to spatial relations, and time and space are apparently inseparable properties of such processes as we associate them with.

In a space-time world a phenomenon escaping space-limitation would

[2] See "The Effect of Distance in ESP Tests," JOURNAL OF PARAPSYCHOLOGY, I, 1937, 172-184.

be presumably time-free. Logically at least the hypothesis may be deduced that the process at work in the distance ESP tests should find no more of a barrier in time itself.

The hypothesis that ESP may be unlimited by temporal conditions thus arises inferentially out of the research already reported. Perhaps to those accustomed to rationalistic solution of problems it may be regarded as quite as strong as the evidence for ESP at a distance. But in science, the empirical solution always takes a more final place than the logical, and fortunately the question here raised is subject to a strictly experimental inquiry.

The general problem of the relation of time and ESP has two aspects, that of reference to future time in the question of precognitive ESP, and coordinately, that of the past in retrocognitive ESP. Greater interest, however, has always attached to the former, to the question of the apprehension of future events rather than of past ones, largely because future ones are of greater practical human import generally. Scientific interest and experimental facility too both support this preference and dwarf the importance of the problem of retrocognition by contrast.

If the order of a pack of cards may be called with an extra-chance degree of success regardless of the exclusion of sensory cues and regardless of distance, it should be expected, therefore, according to the hypothesis stated, that the cards could be called as they will be arranged in some *future time,* with *random rearrangement* intervening. From this reasoning it may be seen that the investigation of the hypothetical ability to precognize the future was a natural experimental development of the ESP research, as well as a logical outcome. The hypothesis was inferable and the methods were readily adapted from the routine ESP test-procedure. To the discussion of these methods I shall shortly return and give further detail.

GENERAL BACKGROUND OF THE HYPOTHESIS

The precognition (or prophecy) hypothesis represents, as is well known, what is for many persons a firm belief, and what has been through the ages an important part of religious and occult creeds and philosophies. There have been too, a great many different beliefs and theories held of the means by which it was believed the future might be read by those possessed of the training or the secret or the gift. These vary widely with different peoples and reflect in general the philosophical and religious backgrounds of the time and place of their origin.

Many such systems of prophecy persist today and enjoy a degree of confidence and extent of practice that are alarming.

The public indulgence in seeking to know the future through these prevailing schemes of supposed prevision indicates at least that there is great interest in the subject, and what is more to the point, that there has not been a convincing scientific study and verdict on the subject. It is true, of course, that scientific thought in general would tend to discourage belief in precognition, but likewise would it at different stages of the history of science have discouraged belief in most of the important subsequent scientific discoveries. It is therefore a question for scientific consideration if only in the interest of social enlightenment and mass mental hygiene. It would surely be healthier for society not to rely upon these false beliefs and deceptive practices if such they prove to be, as would appear most in the line of expectation from the orthodox scientific view of today.

On the other hand, the mere remote possibility of the occurrence of precognition is serious enough to fire the dullest imagination. Its implications are the most far-reaching conceivable, both for the theories of the mind itself and for the view of nature as a whole. There is no question that an adequate scientific demonstration of precognition would produce a major intellectual revolution. This fact alone would suffice to warrant an inquiry into the hypothesis of precognitive ESP.

But there are other grounds that might be given for the pursuit of this hypothesis of prescience. These are represented by the frequently-reported personal experiences of apparent spontaneous prevision of events and by the studies which have been made of these cases.

It is not necessary to the point in question to credit these numerous reports of prophetic insight with the finality of their face value. For the critical student this is impossible. But their collective impression may well lead the thoughtful reader at least to a recognition of the importance of further inquiry into their significance and perhaps to some tentative suggestions as to what to look for and how to go about a systematic research.

With one of the smaller collections of apparently previsionary experiences, largely consisting of dreams of the author himself, that of Mr. J. W. Dunne as reported in his book *An Experiment with Time*,[3] the public is already relatively familiar. Readers can scarcely avoid the impression that there is something to be investigated in the "dreams which come true" as told by Dunne. But it is too easy to pick possible flaws in the method for one to be convinced of precognition. There

[3] London: Faber & Faber, 1934.

is too much latitude for possibly over-favorable interpretation as to success and too much chance for rational judgment to play a part in "predicting" the events of the coming day. What estimate to allow for chance coincidence, we do not know. This and other questions arise in the reading of Dunne's work and prevent acceptance of it at face value. But a residue of conviction remains: There is enough probability to warrant further research!

In fact, Dunne's book led to such a study made along similar lines of dream-recording and daily checking-up with diaries, under Mr. Theodore Besterman of the Society for Psychical Research of London.[4] The results did not bear out Dunne's conclusions, but it should be said that Dunne felt the procedure was somewhat different.

A much less well known but more extensive collection and study of apparently spontaneous precognitive experiences was made by Mr. H. F. Saltmarsh of the Society just named.[5] The cases treated by Mr. Saltmarsh were most carefully selected for authenticity from the large accumulation in the Society's files made over half a century of time, and number several hundred. A statistical analysis for internal consistency lends conviction to the impression of reality given by the report. Saltmarsh himself, in spite of manifest caution, is led to recognize an element of prescience as revealed by the collection. Again, there is the further collection of the late Professor Charles Richet (*L'Avenir et la Prémonition*), and there are other lesser ones.

Considering the frailties of human testimony however, one hesitates to grant so great a point as to feel convinced that the case for prophecy is made. But on the much smaller point by far, that the topic warrants a good looking-into, one is persuaded at once. And if one were not, there is still a further point to make:

Through all the ages there have been no two supposed phenomena more intimately associated than clairvoyance and prophecy. The two, whether true or false, have always gone hand-in-hand. Both in spontaneous experiences and among the professional clairvoyants or prophets have the two been inseparable. The suggestion is inescapable then, that a research which purports to have established the occurrence of the one phenomenon should push on with some justification to a thorough investigation of the other.

This lengthy effort at an explanation of the grounds for taking up

[4] "Report of an Inquiry into Precognitive Dreams," *Proc. Soc. Psych. Res., Lond.*, XLI, 1933, 186-204.

[5] "Report on Cases of Apparent Precognition," *Proc. Soc. Psych. Res., Lond.*, XLII, 1934, 49-103.

the precognition hypothesis is offered because the step will need much justification in the minds of many scientific men of the day. The very fact that an hypothesis of such importance should not have had any systematic experimental investigation prior to the work of the present report is an indication of a deep-seated conviction that to do so is absurd —against all common sense and science.

Indeed, had not the work of Dunne and Saltmarsh freshly raised the question, and the logic of the distance ESP work pointed both to the hypothesis and to an apparently easy method of experimentation, it is very doubtful if the present research could have risen through the resistant layers of indifference and scepticism under which all serious consideration of the question has long lain. In a word, the "days of prophecy" like the "days of miracles" have long been regarded in the halls of science as "past."

PRELIMINARY DEVELOPMENTS

In December, 1933, following his remarkable success in the now well known Pearce-Pratt distance tests, the subject HP was asked to try to predict the order of an ESP pack as it would be after shuffling. He was encouraged to try it by working alone in a preliminary series which would serve to decide only the question of the advisability of further tests.

The procedure was but little different from the established DT procedure for extra-sensory perception. The subject wrote in a record column 25 ESP symbols intending them to coincide with the order which the pack of cards lying before him would have after he had cut it five times in random fashion. The five cuts were intended to represent a very simple preliminary stage of shuffling and thus to introduce a random rearrangement between the predicting and the checking up.

In the 106 runs with this procedure which HP made while working alone he reported an average of 7.1 hits per run of 25 as against the expected chance average of 5 per 25 which was of course a very interesting result. (It is pertinent to recall here that in the 74 runs of distance tests just previously completed he had averaged 7.5 hits per 25. In these distance tests he had no sensory contact whatever with the cards.) It was clearly in order then to attempt this performance under test conditions.

Five other subjects experienced in ESP tests were encouraged to undertake these tests under similar preliminary conditions, except that they were asked to *shuffle* the pack *after* putting the call series on record and before checking up. The shuffling was mainly done by dovetailing

the two halves of the pack together, two such shuffles being given. All but one of these subjects obtained as did HP himself in 116 later runs with this modification, significant scores under the conditions; and the one exception did so later under even better conditions of observation. The totals for each subject are shown in Table I in the left-hand section.

It may be of interest to record that three of the six subjects were graduate assistants in the psychology department at Duke. Their own unwitnessed tests totalled 1,026 runs and averaged 5.8 hits per 25, which gives a critical ratio of 12. The 1,452 total runs of all six subjects averaged 6.0 per run (CR = 19.3). Although these results are not offered as a part of the main experimental evidence, there are some points of secondary interest to which later reference will be made.

The next step was to introduce an observer into the test situation. This was first done with subject HP with myself observing the entire procedure and checking all the records. The observer produced the pack of cards and cut it, and then the subject recorded his 25 calls for the predicted order of the pack. He then took the pack of cards under the observer's full vision and made five random cuts with the pack still inverted. The cards were then turned over one by one by the subject immediately under the observer's eye; and with the double observation, the checkup was made and the score (number of coincidences or hits in the run of 25) recorded by the experimenter.

In 110 runs of this procedure HP averaged 6.1 hits per run, and this gave a positive deviation from the mean chance expectation which represented a critical ratio of 5.7. The performance was evidently of extra-chance nature.

Then HP was given 25 runs in which he was asked to follow the call series with from two to five shufflings of the pack, also in the observer's presence. These runs averaged 6.2 and gave a critical ratio of 2.9. Again it was justifiable to go on to further explorations.

Next a definite period of shuffling (15 seconds) instead of a given number of shuffles was specified, still, however, with continuous witnessing by the observer. In 16 runs the average was 7.7 per run and the critical ratio 5.4.

It was apparent from these series of tests that it did not matter so far as extra-chance scoring was in question whether the cards were cut or shuffled, or just how the shuffling was done.

Three other subjects were introduced to these procedures under observation but with the subject handling the cards between call series and check-up. That is, the subject did the shuffling of the packs. Again two of these subjects were graduate assistants. The work of each

was highly significant, as was that of HP himself. The totals for this set of conditions will be found in the middle section of Table I, giving a grand total of 425 runs, an average per run of 6.1, and a critical ratio of 11.0.

TABLE I.

COMPARISON OF RESULTS IN PRE-SHUFFLING CARD CALLING TESTS UNDER DIFFERENT CONDITIONS OF WITNESSING AND SUBJECT PARTICIPATION

Subject	Method	A. Unwitnessed: Subject Working Alone				B. Witnessed by Experimenter Shuffled by Subject					C. Witnessed and Shuffled by Experimenter. No Contact				
		Runs	Dev.	Av.	CR	Experimenter	Runs	Dev.	Av.	CR	Experimenter	Runs	Dev.	Av	CR
HP...	PDT.	223	469	7.1	15.4	JBR,SOZ,JGP	212	287	6.3	9.7	JBR,SOZ,JGP	482	85	5.2	1.9
GZ...	PDT.	124	21	5.2	0.9	SOZ,MHP....	31	33	6.1	2.9	SOZ.....	654	211	5.3	4.0
SO...	PDT.	260	549	7.1	16.7	GZ,MHP.....	142	93	5.7	3.8	MHP,SOZ....	59	18	5.3	1.1
MHP.	PDT.	642	229	5.4	4.4	.....	....	....	....	....	SOZ.....	66	—6	4.9	0.4
TCC..	PDT.	113	117	6.0	5.4	JBR.....	40	47	6.1	3.7	SOZ,GZ,JBR .	228	—12	4.9	0.4
LHG .	PM..	90	110	6.2	5.7	.....	....	....	....	....	EPG.....	114	43	5.4	2.0
Total.....		1452	1495	6.0	19.1	.....	425	460	6.1	11.0	.....	1603	339	5.2	4.2

*This table is complete for those subjects who performed under at least two of the three general conditions of witnessing. For the complete block of results of the third (right) section, see Table III.

For any problem of less revolutionary consequences these conditions would be adequate to warrant the interest and consideration of critical students of science. But in dealing with the hypothesis of precognition, the slightest loophole would be a fatal defect. Certainly, at any rate, we lose nothing by arbitrarily raising the standards of evidence. For those for whom these preliminary series suffice, they are of course available as evidence. And in fairness to the subjects and the experimenters it must be acknowledged that there is no ground whatever for any active question of the soundness of the tests described above and no reason whatever to believe any possible experimental inadequacy was in any way taken advantage of. (The shuffling of the cards was not screened as in later work because it was regarded as safer to have it under the observer's full attention.)

In the light, however, of the better conditions to be described there is no need to include these preliminary results in the evidence or even to attempt at this point a final estimate of their conclusiveness. We can afford to set aside completely, so far as the main question is concerned, all the tests in which the subject had any tactual contact with the cards during the test.

THE MAIN EXPERIMENTAL SERIES

The principal block of test results which bear on the precognition hypothesis represents something over two-thirds of the total tests made

under the range of conditions to which this report is necessarily restricted. Begun with subject HP early in 1934, this large collection of data has extended over the following four years down to December, 1937, through a total of 113,075 trials aside from the 46,925 trials of preliminary observation referred to above.

The research began in my laboratory under my own observation, but as with the earlier ESP work it quickly attracted the interest and help of others who carried it far beyond my own limited range of achievement. In fact, as in the case of the preceding research referred to, I can only claim to have made a beginning while the main burden of such confirmation and establishment as there has already been was borne by many capable hands that in time took up the exploration.

The names of the experimenters associated with this block of the research are entered opposite the series in which they participated in the table referred to. There are eleven in all. A brief word of introduction may be expected: Three of this group held engineering positions at the time, (Messrs. G. E. Buck, J. L. Michaelson of Schenectady, N. Y., and E. P. Gibson of Grand Rapids, Michigan). Six are already known to readers of ESP literature: (Miss Margaret H. Pegram, Dr. J. G. Pratt, Professor and Mrs. George Zirkle (Sara Ownbey), Mr. J. D. McFarland and myself). All six held at the time positions in a department of psychology, four having been graduate assistants. Dr. Pratt is at present Instructor in Psychology at Duke University and Mr. McFarland holds a similar position at Tarkio College. Mr. Zirkle is Professor of Psychology at Hanover College, and Mrs. Zirkle, sometime Instructor of Psychology at Asheville Normal and Teachers' College. Mr. Earl Stephenson is superintendent of the city school system, Stuart, Iowa; Mr. L. McCartney, an undergraduate student in psychology at Tarkio College.[6]

GENERAL METHODS AND CONDITIONS

The primary condition of this section of the research was that the subject had no tactual experience with the pack of cards on the sequence of which he was making pre-shuffle prediction. The experimenter alone handled them. In much of the work the cards were at a distance from the subject throughout the run but, even when not, obviously there could be no visual cues of a card-order not yet existent. This safeguard is designed to exclude all errors which might be introduced by the subject. If extra-chance results are obtained under these conditions, then

[6] The capable assistance of these contributors to the research is gratefully acknowledged.

it becomes a question only of the care in the handling of the data and the good faith of the experimenters in so far as existing alternative hypotheses are concerned.

Most of the 15 series of tests reported in this paper are based on the calling-before-shuffling modification of the DT procedure, called precognitive DT or PDT. (Exceptions will be noted.) The calls are made and recorded and the pack of cards then shuffled (face down) by the experimenter and checked against the record of the calls. The amount and type of shuffling varied somewhat with different investigators but consisted mainly of the dovetailing method, and of at least two such shuffles.

In making the calls in these PDT tests the subject either wrote down the symbols or called them orally to a recorder. This record in earlier series was checked against the pack by the experimenter (with the subject as a witness) after the "random rearrangement" shuffling was done. In later series a record of the card order was made and the checking done by comparison of call and card records, thus permitting later re-checking.

The exceptions to the test procedure referred to above consisted in adaptation of simple matching techniques to the precognition problem. One of these involved the open matching (OM) procedure. In this modification (POM) the cards were laid face down before a set of five blank spaces instead of the key cards, the key cards to be supplied by chance after the target deck was dealt out. The selection of the key cards afforded the "random rearrangement" since these were chosen by a specified routine procedure from a second pack which was shuffled by the experimenter after the target deck was dealt. The subject of course tried to match the cards dealt against the set of key cards that were yet to come.

In a similar way the screened touch matching (STM) method was adapted to the precognition research (PSTM). The key cards were chosen as just described, and the subject indicated his calls or choices by pointing to one of five empty shallow boxes which *would be expected to have placed in it later* the card to match that which the experimenter had on top of the pack held behind the screen.

In all but one series the experimenter did the checking with the subject also witnessing when this was not prevented by distance. But in the last series, two witnesses were introduced during the shuffling, card recording, and the checking of correspondences in the call and card records.

The subjects who have participated in these tests are all normal

persons. Of the total 49 there is little generalization to be made, except to say that 32 were grade school children in a public school selected from the school at large on the basis of superiority in earlier ESP tests. Nine of the other 14 were students at Duke University who had shown marked ESP ability. One was a professional medium, three were housewives, and one a city engineer.

A portion of the series conducted with the medium was done with the subject in trance, the so-called control personality cooperating (with no difference in results).

One series represented a brief and incomplete comparison of the performance of subjects in a normal waking state with their performance in post-hypnotic period for which supposedly favorable suggestion had been given them while in trance. A fuller report will be given of the series when further research justifies some conclusion as to the possible role of hypnotic suggestion on the abilities under investigation. Only the general results are given here (Table II). The general conditions described above apply also to this series.

Aside from these two exceptions there was no marked peculiarity about the conduct of the 14 test series reported.

All scores obtained under the conditions described are included here.

RESULTS OF THE MAIN EXPERIMENT

As may be seen in Table II, the 113,075 trials given in 15 series to the total of 49 subjects resulted in an average of only 5.14 hits per 25. But over so long a series this gives a deviation of 614 above the mean chance expectation of 22,615. With a standard deviation of ± 137.2 (corrected for the most conservative interpretation) this gives a critical ratio of 4.5 which represents approximate odds of over 400,000 to one that such a deviation would not occur by chance alone. The chance hypothesis is therefore excluded from further consideration.

Detailed data for the individual series may be had from Table II.

DISCUSSION

Low Score Averages: The most disturbing feature of this report, at first inspection, is the fact made apparent in Table I that the more the precautions (in the three general conditions) the lower the scores on the whole. Only two exceptions occur in the six subjects, namely GZ and TCC and these improve with witnessing. From 5.2 to 6.1 for GZ and from 6.0 to 6.1 for TCC. With experimenter shuffling, however, TCC dropped to 4.9 and GZ to 5.3. But it must be said that

TABLE II.

RESULTS OF PRE-SHUFFLING CARD CALLING TESTS IN WHICH THE EXPERIMENTER HANDLED ALL THE SHUFFLING AND CHECKING

Series	Period	Subjects	Experimenters*	Method and Conditions	Runs	Dev.	SD	Av. Score	CR
1....	1934–1935	HP.....	JBR,SOZ,JGP....	PDT mostly at distance.....	482	85	±44.8	5.2	1.9
2....	1934–1935	GZ.....	SOZ,MHP.....	PDT.....	654	211	52.2	5.3	4.0
3....	1934–1935	SO.....	GZ,MHP.....	PDT.....	59	18	15.7	5.3	1.1
4....	1935	HJ.....	SOZ.....	PDT.....	98	17	20.2	5.2	0.8
5....	1935	JGP+EB.....	JGP.....	POM.....	45	12	13.7	5.2	0.9
6....	1935	AJL.....	JBR.....	PDT.....	45	20	13.7	5.5	1.5
7....	1935	GZ,SO,HJ....	GZ,SOZ.....	Pre-hypnotic.....	89	7	19.2	5.1	0.4
				Post " period .	93	56	19.7	5.6	2.8
8....	1935	MHP.....	SOZ.....	PDT.....	66	— 6	16.6	4.9	0.4
9....	1935	TCC.....	SOZ,JBR,GZ.....	PDT.....	228	— 12	30.8	4.9	0.4
10....	1935	EJG.....	JBR.....	PDT.....	948	— 59	62.9	4.9	0.9
11....	1935	MB.....	GEB,JLM.....	PDT.....	35	30	12.1	5.9	2.5
12....	1935–1936	LHG+EPG..	EPG.....	POM.....	154	49	25.3	5.3	1.9
13....	1936	FM.....	JGP.....	PDT,POM,PSTM	1146	41	69.0	5.0	0.6
14....	1937	32 school children....	JDMcF, L.McC,ES.....	PDT.....	335	165	37.3	5.5	4.4
15....	1934–1937	Miscl. 3 Subj..	.....	PDT,POM.....	46	— 20	13.8	4.6	1.4
Total	1934–1937	49 subjects	11 experimenters	PDT,POM,PSTM	4523	614	137.2	5.14	4.5

*The initials of the experimenters represent names as follows: SOZ, Sara Ownbey Zirkle; JGP, J. G. Pratt; MHP, Margaret H. Pegram; GZ, George Zirkle; JBR, J. B. Rhine; GEB, G. E. Buck; JLM, J. L. Michaelson; EPG, E. P. Gibson; JDMcF, J. D. McFarland; LMcC, L. McCartney; ES, Earl Stephenson.

GZ's work for this phase falls into two periods separated by a gap of about a month. The first block of 97 runs average 6.4, the highest he had done for PDT and the second block of 557 runs averaged only 5.1. This later period was a poor one for all test procedures tried.

Likewise this later condition (i.e. experimenter shuffling) coincided with general decline in scoring in the case of subject HP also.

However, the fact remains that the other four subjects declined when the experimenter took over the shuffling. The first thought occurring to the cautious investigator is that this may be due to the additional safeguarding and that conversely the work under the other two sections A and B were experimentally defective. But while this makes little difference since the work of Table II *alone* is offered for consideration in connection with the precognition hypothesis, it should be said in fairness that there are other possible explanations for this decline. Some subjects such as HP believed they were helped by some contact with the cards and liked to be allowed to shuffle the pack in the usual test procedures. (Although HP's best ESP scoring was done without this—with distance and with no sensory card-contact at all.) The most marked exception to this was GZ who also did his best (6.4)

in the 97 first runs at PDT with experimenter shuffling. He wanted no sensory contact with the cards.

Also it is of importance to note that the most advanced condition was tried last and only after considerable preliminary work. The subjects here in the main declined with long series in all lines of procedure.

Even with the decline of scores with the experimenter shuffling, however, the averages are still comparable to the nearest type of test procedure, DT, as Table III shows. And as both Tables I and II indicate, the results obtained under the safest conditions are highly significant anyhow; this, after all, is the really remarkable and puzzling fact, and not why the average is not higher.

The Question of Selection: In this work, *as in all reports from this laboratory,* there has been a description of the conditions of the research and then *a full reporting of all trials and scores made under those conditions.* In Table II series of 20 or more runs are given individual treatment. Three shorter series are grouped under "miscellaneous" but included in the totals.

Hypotheses: Certain hypotheses are ruled out by the evidence, and others have to be considered further. The discussion of these is of first importance.

It was stated above that the results are sufficiently peculiar not to be due to *chance.* It might be said with equally good ground that *sensory cues* were not guiding the subject in making his calls, since sensory cues of future events are not conceivable. Any use either of *rational inference* or of *memory* is excluded by the "random rearrangement" intervening between the calling and the checking.

If the experiments were conducted according to report and the results handled with due care, there is no alternative to the conclusion that extra-chance pre-shuffling calling of card-order has been shown by some subjects. Whether the number of investigations suffices to assure ample confirmation and elimination of possible error and to allay sufficiently any concern as to the experimenter's good faith is a matter for individual judgment. It depends upon the difficulty experienced with the possibility of pre-shuffling calling. The alternative is to await the assurance of further confirmatory reports to follow.

In any case this report adds further strength to the case for extrasensory mode of perception. That is, against the background of the great amount of work already reported, in which the ESP hypothesis has been confirmed, the weight of these further 113,075 trials which are safeguarded even better against sensory cues by the subsequent shuffling is very considerable indeed.

In other words, it may be regarded as easier to apply the evidence of this report to the further confirmation of the long series of reports favoring the ESP hypothesis, while leaving it still difficult to accept it as evidence of precognition. This larger question requires much more evidence even than the former. The only difficulty in this distinction is that in accepting the ESP character of the test results we have no known grounds for not going the rest of the way to a conclusion of precognition. That is, if the results are not due to chance, or sensory cues, or persistent error, or deception, there is no ready-to-hand alternative to the view that ESP is not time-limited.

But in the course of the research there arose new alternative hypotheses (not entirely unheard of but never before seriously considered) which must be disposed of before the present work can be regarded as conclusive evidence of extra-sensory perception of the future. These hypotheses have been themselves subjected to experimental inquiry in this and in other laboratories, and the researches will be reported in following papers.

Of first consideration is the possibility that in the act of shuffling the pack after the predictive calls are made there might be some guidance through ESP itself exercised by the shuffler. That is, through extra-sensory means he may be aware of the card order as well as of the call order (of which he is generally not sensorially aware) to a degree sufficient to stack it somewhat favorably. Only a little favoring is needed, of course, to produce the 614 obtained above the chance mean in 4,523 runs.

We need to know next, then, what the possibilities are for this hypothesis of an "ESP shuffle." Can it be effected by deliberate endeavor by certain subjects? It was necessary to investigate carefully the answer to this question before going on. The preliminary research on this will be reported in the next succeeding number of this JOURNAL. Doubtless it will occur to some readers to suggest that we might avoid the whole difficulty by adopting the use of *mechanical* shufflers. But those who follow the sequence of these reports will discover that this alternative too offers its experimental as well as theoretical difficulties. Suffice it to say here that the question is left unsettled by these and other possible alternatives.

On the other hand it would hardly be fair to the precognition hypothesis to neglect to mention certain evidential features which favor it even against this further alternative of the "ESP shuffle" hypothesis:

First, if the experimenter's shuffling introduced the extra-chance factor, he ought presumably to have success with all the subjects who

call the cards for him. But whereas I had obtained positive scoring as experimenter with subjects HP and AJL, the averages were approximately chance for subjects TCC and EJG. Likewise Mrs. Zirkle, as experimenter, obtained marked positive deviations from subjects GZ and HP but not from MHP and TCC. (However, Mrs. Zirkle obtained high averages from these subjects on other tests as did I also in the instances given above.) Other experimenters had similar differences. It is, of course, remotely possible that a difference could be attributed to particular subject-experimenter combinations even under the "ESP shuffle" hypothesis, and the above argument is far from conclusive.

A second and more urgent consideration lies in the fact that in most instances where comparative data are available, the scores in precognition tests paralleled the course of scoring of the subject on other test procedures. For example, in the instance of subject FM, when her first 20 runs in each of four types of precognition tests[7] were grouped, the average for the 80 runs was 5.71 hits per 25. And when the first 20 runs made by FM on the comparable methods of DT and DTSTM (actually 19 and 24 runs because of irregular natural grouping) were pooled, the total 43 runs average 5.75 hits. The comparison continues through the decline which follows in all six methods: The remaining precognition tests given this subject averaged only 5.0; and for the plain DT and DTSTM, 4.7. This strongly suggests that the subject is following a trend peculiar to herself, and that the shuffling by the experimenter, Dr. Pratt, in the precognition tests was not the determiner of the results.

Again, subjects GZ and HP also underwent a great decline in scoring ability during the period of their precognition tests, although at a much slower rate of decline than FM. This decline appears in both the PDT and in the usual ESP tests and suggests strongly that their scoring ability in PDT was contributed by the subject himself rather than the shuffler of the cards. For example, GZ declined in the BT tests during the period from a score average of 8 per 25 for 46 runs to an average of 6 per 25 for 101 runs. The PDT tests began with 97 runs at 6.4 hits per 25 and after a break fell in the 557 later runs to 5.1 per 25. The case of HP, though it would require more detail to describe, is no less marked in its parallel decline in DT and PDT, all of these cases leave the impression that the three different experimenters involved were not causing the drop in deviation, but that the subject himself actually produced the change.

[7] PDT, POM, PSTM and a special form of POM + STM. See above for account of these procedures.

It would seem wise therefore to regard the above evidence as well as the DT and PDT comparisons below as sufficient to warrant keeping the precognition question well open regardless of the "ESP shuffle" hypothesis, and to suspend judgment in the matter until that hypothesis is further examined in the light of experimental studies. A reasonable position would therefore combine a recognition of the ESP character of the performance tested, with a realization of the possibility that it might not represent actual precognition itself.

Comparison of DT and PDT. If time were to be no restriction upon the capacity for ESP we should not expect to find any appreciable difference between the DT and PDT results yielded by the same subjects under similar conditions. That is, calling before shuffling should be as *successful* as calling afterward (with the pack unbroken, DT). It is worthy of some attention at least that a fairly close approximation is found between the average DT and PDT scores of those subjects having taken tests in both procedures[8] (see Table III below).

TABLE III.

COMPARISON OF SCORE AVERAGES IN TESTS FOR PRECOGNITIVE ABILITY AND COMPARABLE ESP TESTS

Subject*	ESP Tests			Precognition Tests			Remarks
	Procedure	Runs	Av.	Procedure	Runs	Av.	
HP	DT	360	5.3	PDT	482	5.2	Total work of both conditions.
SO	DT	30	5.8	PDT	142	5.7	Not same period. No DT taken in same period as PDT.
HJ	DT	39	5.3	PDT	98	5.2	Approximately same period.
AJL	DT	40	6.0	PDT	45	5.5	Approximately same period.
MHP**	DT	5931	5.3	PDT	642	5.4	Approximately same period.
	DT	1468	4.3	PDT	61	4.3	Approximately same period. (Low-aim series, both DT and PDT).
EJG	DT	152	5.3	PDT	948	4.9	Two methods used—a year apart.
MB	DT	24	6.6	PDT	35	5.9	
LHG	OM	96	7.0	POM	114	5.4	Different periods Precognitive matching used.
	DT	46	5.6				
FM	DT	43	5.7	PDT	80	5.7	Approximately 1st 20 runs of each of 2 DT methods and of 4 precognition test methods used with this subject.
	DT			PDT			
	DT STM	200	4.7	POM	1066	5.0	Remainder of tests, both conditions.
School children	DT	218	5.3	PDT	335	5.5	

*Subjects of Table II omitted from this table are those who did not perform on the ESP test most comparable; hence there is no basis for comparison. The school children in No. 10 series were not the same since both sets of tests, DT and PDT, were not given either group. But both groups are from the same school, are approximately of the same age and status, and were tested by the same experimenters.
**Unwitnessed work, but this subject was a psychology assistant.

[8] In the case of subject LHG the methods were OM and POM and this also shows the greatest difference of any series in the table. This subject's DT average is entered for comparison.

A general inspection of this table will undoubtedly leave the impression upon those familiar with the score ranges obtained in other ESP tests (varying from below 3 hits per 25 to above 18) that there is some unusual similarity in these DT and PDT scores—that the subject's ability in PDT is essentially that of DT scoring, and that the temporal factor is not inhibiting. But again, it could not be regarded as more than strongly suggestive of the occurrence of precognition.

The research may be credited, however, with having produced some considerable presumptive probability of the occurrence of precognitive ESP and then to have raised a new alternative hypothesis which points to a new line of experimentation. Whichever hypothesis proves to be correct, precognition or the "ESP shuffle," a new advance may properly be claimed for the research.

SUMMARY

There has been reported, above, the work of eleven investigators in testing 49 subjects for ability to call cards in advance of the supposedly random rearrangement of simple shuffling by the experimenters. In the 4,523 runs a total score of 614 above mean chance expectation was obtained, which is 4.5 times the standard deviation. The odds are over 400,000 to one that so great a deviation would not be expected to occur by chance alone.

With the chance hypothesis ruled out, with sensory cues impossible from an arrangement of the cards not yet in existence, and the shuffling safeguarded and handled entirely by the experimenter who also did all the checking, there is left no serious alternative to the precognitive hypothesis except that of conceivable unreliability of research personnel or the possibility (not an established one) of the experimenter unwittingly shuffling the pack, guided by ESP, so as to match to a significant degree the predictive calls made by the subject. The first alternative will have to be weighed non-statistically by the individual readers, whose criteria for tentative acceptance will vary greatly. The second hypothesis, the "ESP shuffle," is the subject of another research and a succeeding report.

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[14]

Methodological Criticisms of Parapsychology

Charles Akers

1. Introduction

In his presidential address to the Parapsychological Association, Rao (1978) claimed that the evidence for psi was "inescapable" and that criticisms of the field were "unfair" and "false" (p. 278). Yet the field of parapsychology has come under increasing criticism in recent years. Some of the sharpest criticism has come from Fellows or Consultants of the Committee for the Scientific Investigation of Claims of the Paranormal. Hansel, one of the Fellows, has written a revision (1980) of his 1966 work, which has been so widely cited. Other volumes of criticism have been contributed by Alcock (1981), Gardner (1981), Marks and Kammann (1980), and Randi (1980). Frazier (1981) has edited a collection of articles from the *Skeptical Inquirer*, many of which deal with parapsychology. Psychologist Hyman (1976, 1977a,b, 1981a,b, 1983) has written a number of serious reviews of parapsychological literature.

The recent spate of criticism, greater than at any time since the 1930s, has not been confined to the Committee. Girden (1978) has updated his 1962 review of psychokinesis to cover parapsychology as a whole. Diaconis (1978), Moss and Butler (1978), and Zusne and Jones (1982) have published other major critiques. There have been many important articles and volumes of criticism, too numerous to cite here.

The response by parapsychologists to this barrage of criticism cannot be readily summarized. There are serious disagreements within the field over the strength of the evidence for psi (McConnell & Clark, 1980; Truzzi, 1980). However, there does seem to be agreement that the critics in general, and especially Hansel, have wrongly focused on the results from "critical experiments" or from studies with famous "psychics." There has, for example, been an enormous amount written on psychic showman Uri Geller, whom Morris (1980b) refers to as "a tired old favorite of the critics" (p. 435). Honorton (1981), Morris (1980b), Palmer (1978), Rao (1980) and other parapsychologists have argued that the case for psi depends not on isolated experiments with selected psychics, but on repeatable experiments with more-or-less unselected subjects. Claims of repeatable findings are discussed in the Wolman (1977) *Handbook of Parapsychology* (especially in chapters by Honorton and Palmer) and in volumes of the present series.

This response is a fair assessment of current criticism; with a few exceptions (e.g., Hyman, 1983; Marks, 1982) the critics have dealt with particular experiments, and have not attempted generalizations about an entire area of research. Any review of published criticism will, then, also fail to make generalizations. As a partial remedy for the situation, I have supplemented the present review with a survey of 54 ESP experiments, which have been cited in the Wolman *Handbook* or elsewhere. I included in the sample 32 experiments where "psi-conducive" procedures were used, as discussed by Honorton (1977), and another 22 experiments where ESP scores were significantly correlated with personality or attitude variables (see Palmer, 1971, 1977, 1978, 1982). These are the areas of research where the repeatability of results seems to be the best. The question is, to what extent are the sample studies contaminated by procedural flaws?

Results from the sample will be prefaced by general discussions of methodological issues, which are introduced under the following headings: Randomization Failures, Sensory Leakage, Subject Cheating, Recording Errors, Classification and Scoring Errors, Statistical Violations, Reporting Failures, and Experimenter Fraud. In each section I will discuss not just experiments from the survey, but also other studies where these issues have been raised. I will not attempt to cover all published criticisms, but will focus instead on those that have the most relevance to mainstream parapsychological research.

2. Selection of the Sample Experiments

In designing the sample, I chose experiments from areas of research reviewed by Honorton (1977) and Palmer (1977, 1978). Honorton has reviewed altered states of consciousness research, while Palmer's reviews are of personality or attitude correlates. I confined myself to studies where significant results had been claimed for a sample of more-or-less unselected subjects ($N \geqslant 5$). I did not include in my survey any studies of highly selected subjects, for two reasons: First, such subjects cannot usually be retested, either because they have lost their abilities, or because they are unavailable. Second, these special subjects (such as Stepanek) tend to have specialized skills, the demonstration of which involves methodological questions which do not necessarily have much relevance to ordinary parapsychological research.

I did want the sample to reflect serious research efforts which had warranted a full publication. For this reason I chose to exclude all reports which were unpublished, or published merely as abstracts, convention reports, etc. I also chose to exclude weak experiments which were preliminary to a stronger experiment published within the same report (e.g., Casler's 1962 preliminary work, or Fahler and Cadoret's 1958 Exps. A & B). Two interesting experiments by Honorton (1964, 1966), cited in Honorton's (1977) review, were excluded only because the hypothesis was a complex one which has not yet been confirmed by other investigators.

At this point I was left with 42 experiments which either Honorton or Palmer

had cited as significant confirmations (Honorton using an .05-level one-tailed criterion, and Palmer an .05 two-tailed criterion). These were all ESP rather than PK experiments. To this group I added 12 more recent (through 1982) experiments from the same research areas. I excluded a number of studies which had achieved some significant results, but did not appear to have met the criteria set by Honorton or Palmer (e.g., Casler, 1976; Kesner & Morris, 1978; Rao, Dukhan, & Rao, 1978, Drawing Exp.; Shrager, 1978; Sondow, Braud, & Barker, 1982; Stanford, 1979). In line with Honorton's (1977) criteria, I looked for significant scoring in an altered state condition, irrespective of whether scoring in that condition differed from a control condition. I did not, however, include any studies where the scores were in the unanticipated direction (e.g., Child & Levi, 1979).

The final sample of 54 experiments is fairly complete. If it is not inclusive, it is at least representative of findings in altered states and personality research. The composition was as follows:

1. Eleven ganzfeld experiments (Braud & Wood, 1977; Braud, Wood, & Braud, 1975; Honorton & Harper, 1974; Sargent, 1980b, Exps. 2, 3, & 5; Sargent, Bartlett, & Moss, 1982; Schmitt & Stanford, 1978; Sondow, 1979; Terry & Honorton, 1976, Exps. 1 & 2).
2. Twelve hypnosis experiments (Braud & Mellen, 1979; Casler, 1962, 1964; Fahler & Cadoret, 1958, Exp. C; Grela, 1945; Honorton, 1972; Honorton & Stump, 1969; Krippner, 1968; McBain, Fox, Kimura, Nakanishi, & Tirado, 1970; Moss, Paulson, Chang, & Levitt, 1970; Parker & Beloff, 1970, Exp. 1; Sargent, 1978a).
3. Twelve studies of personality correlates (Casper, 1952; Haraldsson, 1978; Johnson & Kanthamani, 1967, Exp. 1; Kanthamani & Rao, 1971, Pilot, Exps. A, B, & C; Nash & Nash, 1961; Nicol & Humphrey, 1953; Sargent & Harley, 1981; Shields, 1962, Exps. 1 & 2).
4. Ten studies of attitude correlates (Bevan, 1947; Bhadra, 1966; Carpenter, 1971, Exp. 2; Moss & Gengerelli, 1968; Musso, 1965; Ryzl, 1968; Schmeidler, 1971, Pilot; Schmeidler & McConnell, 1958, Individual and Group Exps.; Wilson, 1964, Exp. 1).
5. Five relaxation experiments (Braud & Braud, 1973, Exp. 2; Braud & Braud, 1974, Exps. 1 & 2; Braud & Braud, 1977; Stanford & Mayer, 1974).
6. Four meditation experiments (Rao, Dukhan, & Rao, 1978, Pilot, Exps. 1 & 2; Schmeidler, 1970).

The reader will note the inclusion of three studies designated as "pilot." These were included in the sample because their methods were similar to those in later studies designated as "confirmatory" or "follow-up." In general, designations such as "pilot," "exploratory," or "preliminary" were unreliable guides to the actual rigor with which an experiment had been carried out. Often, a study is "preliminary" with respect to some hypotheses and "confirmatory" with respect to others. Some studies designated as "exploratory" were among the best in the sample, from a methodological standpoint. Others, designated as "confirmatory," were among the weakest in the sample.

3. Randomization Failures

3.1. Introduction

A weakness in a number of "classic" ESP experiments was the employment of target sequences that were severely nonrandom. One such case was the Groningen telepathy series, where the targets were letter/number combinations. Schouten and Kelly (1978) examined the sequences, and found that the target letters were not equally frequent. Letters (A through H) were distributed as follows: 103, 102, 74, 43, 73, 71, 98, 23, with $\chi^2(7) = 78.6$, $p < .001$. There were similar inequalities in the distribution of target numbers, and in the letter/number combinations. Schouten and Kelly claimed that the nonrandom features could not account for the ESP scoring. Whether or not a nonrandom feature can "account for" an ESP effect, it is a disquieting situation when the departures from randomness are severe, and of unknown origin. If a critical procedure such as randomization broke down, this raises the possibility that there were other breakdowns, or failures to carry out the experimental plan. Each such failure lessens one's confidence in the competence of the investigation as a whole. In the usual ESP investigation, only one peculiarity of the results should emerge: a deviation from the number of "hits" predicted by probability theory.

Nonrandomness was also found in the Reiss series, regarded by Pratt, Rhine, Smith, Stuart, & Greenwood (1940/1966) as one of six conclusive ESP experiments. The target sequence from the series was deficient in target doubles (Nicol, 1959). This deficiency could not account for the scores achieved, but the origin of the nonrandom feature was again unknown. Also discussed by Nicol (1959) was a highly significant deficiency of certain triplet combinations in target sequences from the classic Soal-Goldney experiments. The targets had supposedly been derived from tables of logarithms, but the source could not be verified. Evidence eventually emerged that Soal, the chief experimenter, may have altered some targets to create false extra hits (Hansel, 1980; Markwick, 1978; Scott & Haskell, 1974). The evidence is strong, though circumstantial. In this case, the nonrandom sequences were a clue to experimenter error. This possibility must always be considered in cases where randomization breaks down.

3.2. Randomization by Shuffling or Other Informal Techniques

When target sequences are obtained by manual shuffling of card decks, deviations from randomness can always be excused in terms of faulty shuffling techniques, or faulty cards (e.g., Schouten & Kenny, 1978). Wilson (1966) saw no reason for the use of card decks, and argued that "in future all ESP tests should be carried out with series of random numbers which we produce by a clearly understood physical process, so that the limits of their accuracy may be clearly

stated" (p. 393). Davis and Akers (1974) came to a similar conclusion but were more open towards the use of informal methods in preliminary research. Yet, shuffling and other informal techniques of randomization are often used in research beyond the preliminary phase. Stokes (1977) criticized Targ and Puthoff's (1974) experiments on this basis. On some trials the investigators had selected targets for the subject, Uri Geller, by "randomly" opening a dictionary and interpreting the first drawable word in whatever manner they saw fit. The use of such a procedure raises questions about the extent to which the Geller experiments were preplanned. If the experiments were not preplanned, then there may indeed be grounds for suspecting reporting errors, as Marks and Kammann (1980) allege.

The Targ and Puthoff research is not an isolated example. Hyman (1983) surveyed 42 ganzfeld ESP experiments (most of which had not been fully published), and found numerous failures to describe randomization procedures. Where the procedures were described, they were often informal (e.g., shuffling). Often, researchers mentioned the use of "random numbers," but with no account of where the numbers came from, or how they were employed (personal communication, Hyman, 1983).

3.3. Randomization in the Sample

Among the 54 experiments in my sample, randomization was by shuffling or other informal techniques in 20 cases (Bevan, 1947; Carpenter, 1971, Exp. 2; Casler, 1962, 1964; Casper, 1952; Fahler & Cadoret, 1958, Exp. C; Grela, 1945; Honorton, 1972; Honorton & Stump, 1969; Johnson & Kanthamani, 1967, Exp. 1; Kanthamani & Rao, 1971, Pilot, Exps. A, B, & C; Parker & Beloff, 1970, Exp. 1; Rao, Dukhan, & Rao, 1978, Pilot, Exps. 1 & 2; Shields, 1962, Exp 1; Sondow, 1979). In another six cases randomization was incompletely described (Honorton & Harper, 1974; Terry & Honorton, 1976, Exps. 1 & 2), or given no description at all (Moss & Gengerelli, 1968; Moss et al., 1970; Shields, 1962, Exp. 2). All told, randomization was by informal means, or was poorly described, in nearly half the sample (26/54). Among the remaining 28 experiments, randomization was by random number tables in 27 cases, and by ESP test machine in one case (Haraldsson, 1978).

The use of shuffling or similar methods denotes a casualness in the design of what Honorton (1977) or Palmer (1977, 1978, 1982) accept as confirmatory studies. Nonetheless, there is an argument for considering this as a "minor contaminant" rather than a "major flaw," in most such cases. The main concern is that subjects will learn the characteristics of previous target sequences, and use this information to infer the nature of future target sequences. To do this, they need periodic feedback on the targets. Yet, such feedback was not available in four of the 20 questionable experiments (Casler, 1964; Fahler & Cadoret, 1958, Exp. C; Honorton, 1972; Honorton & Stump, 1969). In two other studies, by Parker and Beloff (1970, Exp. 1) and Sondow (1979), the strongest results were

from sessions where no feedback was given. This leaves 14 experiments where feedback effects were a hypothetical possibility. Of these experiments, eight were "blind matching" card tests, in which subjects place cards from a "target deck" beneath whichever of several "key cards" they think will yield a match. Presumably, neither target cards nor key cards are known to the subjects. To succeed at the task through inferential strategies, subjects would need to remember the orders of both target cards and key cards from the previous run, and use this information to predict their future joint distribution. This is a rather involved counterhypothesis, which I cannot consider except as a remote possibility (with these unselected subjects).

If the blind matching studies are not considered as seriously flawed, this still leaves six experiments where inferential strategies would seem to be a real possibility (Bevan, 1947; Casler, 1962; Casper, 1952; Grela, 1945; Johnson & Kanthamani, 1967, Exp. 1; Shields, 1962, Exp. 1). With the exception of Casler (1962) these studies appear to be flawed (pending study of the actual sequences of calls and targets). Casler (1962) used four different decks, and feedback was given only after four runs, so I chose not to assign a flaw (that is, to consider the experiment flawed in this respect). In the remaining cases there was feedback after each run (Bevan, 1947; Casper, 1952; Grela, 1945; Shields, 1962, Exp. 1) or after each trial (Johnson & Kanthamani, 1967, Exp. 1). In none of these cases was it reported that the target deck was changed between runs. If a single deck was used repeatedly, with run-by-run feedback, then subjects may well have discovered a guessing strategy. This is even more likely, as Diaconis (1978) has argued, where subjects obtain trial-by-trial feedback. However, that situation existed only in the study by Johnson and Kanthamani.

It could be argued that informal randomization invalidates any ESP experiment. Even if subjects are kept from seeing past target orders, their calling preferences might match patterns in the target sequence in such a way that the theoretical variance of ESP scores increases (Feller, 1940). If this happened, the usual statistical procedures would be invalidated. I would agree that this is a theoretical possibility, but I know of not a single ESP study (excluding the multiple-calling situation) where such an increase in variance has actually been observed. If unbalanced decks are used, as in the examples provided by Zusne and Jones (1982), then psi artifacts can easily emerge. However, all the card experiments cited above involved balanced decks.

There were 27 cases in the sample where random number tables had been employed, and one case where an ESP test machine was used (Haraldsson, 1978). Since there was no description of the test machine, nor any report of randomness tests, the experiment by Haraldsson must be considered flawed, until the situation can be clarified. In the cases where random numbers were used, it was not always clear exactly how a target sequence was finally generated. However, the application of random number tables, to obtain an open-deck target sequence, is a procedure that is familiar to most parapsychologists. In one study by Braud and Braud (1973, Exp. 2) the tables were applied by untrained agents, and a misapplication is possible; the study appears to be flawed in this respect.

There were other studies (e.g., Krippner, 1968) where the target selection procedures were assigned to experimental assistants, whose degree of training and/or supervision is uncertain. However, I did not assign a flaw in those cases.

In summary, there were 20 cases in the sample where informal randomization was used, and six cases where randomization was poorly described. Five of the former cases were considered flawed, as were all of the latter. Flaws were also assigned to studies by Braud and Braud (1973, Exp. 2) and Haraldsson (1978), bringing the total number of such experiments to 13 out of the 54. This may well understate the problem, the extent of which can only be known by checking the randomness of the target sequences (Davis & Akers, 1974). This was done in only one instance (Schmeidler & McConnell, Group Exps.). When Schmeidler (1959) checked her sequences, she found that an assistant had inadvertently used many closed-deck target orders, rather than the open-deck orders that she had requested. Fortunately, this was an error that required only minor revisions in her data analyses. One wonders whether similar errors may have occurred in other experiments from the sample, with more serious consequences.

3.4. Randomization by ESP Test Machine

The study by Haraldsson (1978), discussed above, was the only case within my sample where randomization was by ESP test machine. This was a fixed-choice ESP experiment, whereas most of the recent studies in the sample were free-response experiments, where each subject contributes only a few trials (often just one). In such small-scale experiments there is no need for a test machine.

Schmidt (1969b, 1970b) devised the first fully automated test machines, and obtained highly significant results with them (Schmidt, 1969a,b). These early experiments were critiqued by Hansel (1980), who claimed to have found numerous sources of error (see Section 7.1), some of which concerned possibilities for machine bias. Although Hansel claimed no evidence of machine bias (Schmidt's devices had passed all randomness tests) he found the design of the machine defective, in that there was no mechanism for correcting bias if it did arise. According to Hansel, Schmidt ought to have used a "switching system" (p. 231). With the use of Hansel's "switching system," the possibility for systematic bias is eliminated. I believe that Hansel's suggestion is a useful one. It should be possible to devise a system which would control not only for first-order bias, but also for higher-order sequential dependencies (e.g., a patterning in the target sequence). Schmidt, at a 1974 research meeting at the Foundation for Research on the Nature of Man, suggested doing this by incorporating the Rand Corporation (1955) tables into such a switching system.

If the test is for clairvoyance rather than precognition, then the Rand tables can be used directly (or with mixing) as the machine's target source. Schmidt (1969a) conducted such tests, and again obtained highly significant ESP scores from his subjects. Hansel's (1980) critique of these machine tests focused on

possibilities for experimenter error, especially in the assignment of scores to "high-aim" and "low-aim" conditions (see Section 7.1).

When targets are obtained from a random event generator (REG) rather than from a predetermined source, then it is critically important to include control runs. Schmidt (1969b) had introduced such runs, but not in a systematic fashion. Both Hansel (1981) and Hyman (1981a) have advocated the systematic introduction of control runs (as in a counterbalanced sequence of control and experimental runs). As Hyman (1981a) notes, the REGs are continually being modified or replaced. There is not much standardization, which makes it all the more important to introduce systematic controls, especially as a guard against short-term generator bias.

3.5. Tart's Ten-Choice Trainer

The need for control runs is especially apparent in some initial test machine experiments by Tart (1975). Tart had used several machines, of a different design from Schmidt's, in three "ESP training" studies (Tart, 1975, 1981a; Tart, Palmer, & Redington, 1979). The methodology of the first study was severely criticized both within parapsychology (Kennedy, 1980a,b; Stanford, 1977c,d) and without (Gardner, 1977, 1981; Gatlin, 1978a,b, 1979). Readers who wish to make a full appraisal of Tart's "ESP training" studies should consult the critiques cited, and Tart's replies (Tart, 1977a,b, 1978, 1979a,b, 1980, 1981b; Tart & Dronek, 1980). I will deal with only the first training study, and confine myself to the question of whether nonrandom target sequences might explain the direct hit scoring. There is an important issue involved, which justifies detailed discussion: To what extent can one make allowances for nonrandom target sequences, and salvage an experiment which is flawed in this respect?

In Tart's (1975, Chapter 4) first training study, he used two test machines, an "Aquarius Model 100" and a "Ten-Choice Trainer." The Aquarius machine had been tested before the study began, and also at its end, by generating a control sequence of 1,000 trials. Tests on the sequences indicated no significant departures from randomness. However, as Stanford (1977d) observed, these tests were limited and did not include any desired checks on higher-order sequential dependencies (analyses beyond the doublet level). In the midst of the training phase the Aquarius broke down and began producing nonrandom sequences (Tart, 1975, p. 163). It then became necessary to discard some data and have the machine repaired before the study could proceed. Were some data retained that ought to have been discarded? Or might the Aquarius have malfunctioned in ways that escaped the experimenters' attention? These questions cannot be answered, since Tart failed to arrange for a trial-by-trial record of Aquarius data (Tart, 1977c, p. 89); there can be no study of the target sequence. Such a failure to record the target sequence is a severe shortcoming in any ESP experiment.

Tart did provide for a trial-by-trial record of data from the Ten-Choice Trainer. However, when this record was examined, highly significant departures from

randomness were found (Gatlin, 1979; Goldman, Stein, & Weiner, 1977; Kennedy, 1980b). Goldman, Stein, and Weiner (1977) concluded that "until the experiment is done again, we are in the position of a chemist who at the end of an experiment discovers that his test tube was dirty" (p. 37).

In reply to such criticism Tart (1977a, 1978, 1979a, 1980) acknowledged the deviations from randomness, but saw these as small relative to the presumed ESP effects. He was able to explain the initial finding by Goldman, Stein, and Weiner (of a deficiency in target repetitions) in terms of an innocent error by his undergraduate experimenters (Tart, 1980, p. 217). Tart (1980) speculated that other findings of nonrandomness, for which he did not have a ready explanation, could have arisen from a PK influence on the test machine.

Kennedy (1980a,b) agreed that PK was a possible explanation for the nonrandom features. However, he saw experimenter error as an alternative explanation. Kennedy pointed out a fundamental weakness in the procedure: The Ten-Choice Trainer was not completely automated, since the experimenter had to manually enter targets by setting a switch. If the experimenter entered a target other than that chosen by the REG, then nonrandom artifacts could result. Kennedy stressed the fact that only one experimenter out of six had achieved significant results with the Ten-Choice Trainer, and the nonrandom effects were strongest in his data. Hence, there was no need to assume ESP. The scoring might have arisen from errors by the experimenters (e.g., entering a target which the subject was about to choose). If there were no such errors, it was still the case that the targets were nonrandom. Hence the subjects, who were highly selected, might have used guessing strategies, rather than ESP.

For purposes of the present discussion, I will focus on the possibilities for guessing strategies (Tart's subjects did have trial-by-trial feedback). To answer this criticism Tart and Dronek (1980) developed a "Probabilistic Predictor Program." This was a computer program which mimicked ESP by utilizing all information from previous targets to predict the next target in the sequence. Despite its advantage of perfect memory, the computer program was unable to match the success of Tart's ESP subjects. On this basis Tart concluded that only a small portion of the ESP scoring could be explained by inferential strategies.

At first sight, Tart's approach seems reasonable. As Wilson (1966) has argued, all ESP test machines will exhibit small biases. Hence, it is always a matter of taking bias into account when evaluating an ESP effect, and this is what Tart has tried to do. However, Wilson was referring to cases where (a) the bias was a more-or-less unvarying characteristic of the machine, and (b) the source of the bias was well understood. Neither of these criteria were met in the case of the Ten-Choice Trainer. Moreover, the biases in the target sequences obtained from the Trainer, in the first training study, were much larger than those envisioned by Wilson for a well-designed machine. The biases were large enough that Tart's highly selected subjects may have taken advantage of them, even though his computer program could not do so very effectively. There are still many human talents that computers cannot match. Otherwise, we should conclude that a person is psychic whenever he or she defeats a computer program.

Tart (1980) regards the target biases as "slight deviations from randomicity" which "have no real empirical consequences" (p. 219). This is a recurrent theme in the psi controversy: Where critics see a "dirty test tube," the original investigator sees a minor contaminant. In the first training study the contaminant was a nonrandom target sequence. However, none of Tart's critics were able to show that this could account for the extrachance scores achieved by his subjects. Since the scores remained unexplained, Tart felt free to attribute them to ESP. In doing so he made a shift from the usual logic of an ESP experiment. The usual logic is to suppose an ESP effect only when all other alternatives have been ruled out—not by post hoc analyses, but by the design of the experiment. The reason for this is that ESP is not an established process; it is only a label that we assign to certain anomalous findings. To verify the anomaly it is necessary to absolutely rule out sources of error, as Humphreys (1968) has argued. Humphreys' approach is the conservative one adopted in this chapter.

3.6. Randomization in PK Experiments

Early PK studies, begun by Rhine in 1934, usually involved dice (thrown by hand, cup, or machine). Girden (1962) was severely critical of this early research, and particularly emphasized the lack of controls against dice bias. Stanford (1977b) suspects that Rhine recognized the weakness and delayed publication on this basis (for nine years!). When the results were published, Rhine emphasized not the overall scoring, but a decline of scores across the record page, which could not readily be attributed to dice bias (as in Rhine & Humphrey, 1944). However, there were weaknesses in the research other than dice bias, and Stanford surmised that "spurious interactions might, in fact, occur in a number of ways having nothing to do with psi" (p. 326).

In later (post-1944) studies, dice bias was usually controlled for by designs in which the subjects tried for all die faces an equal number of times (e.g., McConnell, Snowdon, & Powell, 1955). However, such studies did not include a "nonwishing control" group, of subjects who did not desire any particular outcome. Girden (1962) saw the lack of such a control group as a fundamental weakness. Murphy (1962) objected that there was no need for such a control. It was already the case that subjects, while wishing for one face, were not wishing for the other five faces. Murphy saw this "internal control" as sufficient. It seems to me that Murphy had a good argument. (See Section 8.1 for a further discussion of this issue.)

In modern PK experiments, the outputs of the REGs are sometimes switched during the course of an experiment; this is analogous to switching the hoped-for target face in the dice studies. Since the switch is usually made only once during the experiment, it is not a very effective control against generator bias (which might change over time). More commonly, researchers rely on the introduction of control runs. In the first fully automated PK experiment, Schmidt (1970a) included extensive control runs. However, as Hansel (1981) observed, these

were not introduced in any systematic fashion. Hyman (1981a) agreed with Hansel, that this was a weakness in the Schmidt experiment, but he did not see it as a fundamental failure. I am inclined to agree with Hyman. Schmidt did, at least, conduct extensive control runs at the time of the experiment.

Hyman (1981a) observed that PK research by Schmidt and others was still mired in its preliminary phases, since (a) the generators were "neither standardized nor debugged," and (b) results were "far from lawful, systematic, or independently replicable" (p. 39). A survey of 214 PK experiments, conducted by May, Humphrey, and Hubbard (1980) appears to substantiate Hyman's claim. The authors found that none of the experiments had been adequately designed and written up. Their four summary points are reproduced below:

> (1) No control tests were reported in more than 44% of the references. Of those that did, most did not check for temporal stability of the random sources during the course of the experiment.
> (2) There were insufficient details about the physics and constructed parameters of the experimental apparatus to assess the possibility of environmental influences.
> (3) The raw data were not saved for later and independent analysis in virtually any of the experiments.
> (4) None of the experiments reported controlled and limited access to the experimental apparatus [p. 8].

As a check on the authors' first conclusion I examined 27 PK experiments cited by Honorton (1978b) which had (a) achieved significant results (by Honorton's criteria), and (b) been published in full. In only 17 of the 27 were randomness tests reported at the time of the experiment, and, as claimed by May, Humphrey, and Hubbard, these tests were for overall bias, and did not (with the exception of the study by Schmidt, 1970a) ensure detection of short-term generator bias. The 17 experiments were all authored (or coauthored) by Schmidt or Braud, except for one by Bierman and Hootkooper (1975-1977). In none of the studies were randomness tests introduced in the systematic fashion advocated by critics, though this has been the case in more recent research (e.g., Bierman & Houtkooper, 1981; Broughton, 1979, 1982). It was generally unclear whether the REGs had been tested in the actual experimental environment, with all peripheral equipment attached. This is a fundamental requirement for a randomness test (Davis & Akers, 1974).

One means of systematically controlling for generator bias is to randomly pair control and experimental trials, as suggested by Hansel (1981). This is a procedure that Broughton began routinely applying several years ago (e.g., Broughton, 1979), and it is quite easily accomplished when one has an REG interfaced to a computer (Broughton, 1982). However, this is not to imply that there are no other means of accomplishing such control. Where several levels of a "psi-conducive variable" are manipulated, the sequence of conditions could be determined by randomly selecting latin square patterns (Cochran & Cox, 1957). However, this would have to be carefully planned out, since one needs to control not just for some linear trend in the generator, but also for short-term biases,

and sequential dependencies. If a single latin square pattern was used repeatedly, it might coincide with a pattern that the generator exhibited. It should be mentioned, however, that some Schmidt generators (e.g., Davis & Akers, 1974, pp. 403-406) have been tested by generating sequences of over a million trials, and have shown no evidence of either short- or long-term bias. Hence, the problem is not a severe one with a well-designed generator which has been thoroughly tested.

4. Sensory Leakage

4.1. Introduction

In an ESP experiment the design should exclude cues from the target itself, from an agent who "sends" the target (in a telepathy test), or from the experimenters who have prepared the target. If judges are employed, as in many free-response ESP tests, the design must also exclude leakage to them. These issues will be discussed in turn.

4.2. Leakage from the Target in Clairvoyance Tests

J.B. Rhine's (1934/1964) initial card experiments were severely criticized for failures to exclude sensory cues (Hansel, 1980; Herbert, 1938; Kennedy, 1938). Cards were frequently used under unscreened conditions, and where screens were used, these were often just the experimenter's hand or notebook. The early cards, which were cut from heavy cardboard, must have been difficult to shuffle, especially since they were not always cut to the same size (Rhine, 1934, p. 152). Rhine's most famous subject, Hubert Pearce, was typically allowed to shuffle the cards himself, though Rhine or another "observer" would give them a final cut. Pearce did succeed in one experiment (Rhine & Pratt, 1954) where the cards were at a distance, but there may have been opportunities for cheating (see Hansel, 1980; Honorton, 1981).

In later work at Duke, a controversy developed when it was discovered that some commercially cut ESP cards could be read from their backs (Herbert, 1938; Kennedy, 1938). After this Rhine began advising against any use of unscreened cards (see "Notes" in the December 1937 issue of *Journal of Parapsychology*). Standard methods of screening evolved, as in the Pratt-Woodruff experiment (Pratt & Woodruff, 1939).

However, Schmeidler (1977a) makes no mention of screening techniques in the Wolman *Handbook of Parapsychology*. Instead, she advises that "if targets are near to the subject, they must be in an opaque container, unopened until the subject's responses have been recorded" (p. 133). I would add a proviso that the opaqueness of the container be objectively assessed (see Marks & Kammann,

1977) and that the container be kept well out of the subject's reach. An alternative would be to allow the subject unlimited access to the container, which was made fraudproof. Price (1955) suggested that a metal container be used with a cover welded on and photomicrographs taken of the welds. But perhaps such a container could be accessed today by the same technology that is used to examine oil pipelines (advanced imaging techniques). Rather than attempt to devise the perfect container (which would in any case be impractical for day-to-day research), an experimenter can simply take precautions to ensure that the subject has no visual, tactile, or other access to the target container at any time during the experiment. This may not be as easy as it sounds, if the target must be prepared in advance, and the subject is a clever trickster.

The same principle, of limiting the subject's access, applies to computer storage of targets. Davis (1974) has observed that a subject who has access to the computer, knows the data format, and has sufficient programming knowledge, might subvert experimental precautions. Davis also noted that software "bugs" might allow easy access to computer targets, in a system which was supposedly secure.

4.3. Target Leakage in the Sample

There were 33 clairvoyance experiments in my sample, and among these there were 17 cases where the targets were concealed somewhere within arm's reach of the subject (Braud & Braud, 1977; Braud & Mellen, 1979; Carpenter, 1971, Exp. 2; Honorton, 1972; Honorton & Stump, 1969; Johnson & Kanthamani, 1967, Exp. 1; Kanthamani & Rao, 1971, Pilot, Exps. A, B, & C; Parker & Beloff, 1970, Exp. 1; Rao, Dukhan, & Rao, 1978, Pilot, Exps. 1 & 2; Schmeidler, 1971, Pilot; Schmitt & Stanford, 1978; Shields, 1962, Exp. 2).

In another five experiments it was unclear what distance separated subject and target (Casler, 1962; Fahler & Cadoret, 1958, Exp. C; Nash & Nash, 1961; Shields, 1962, Exp. 1; Wilson, 1964, Exp. 1). This leaves 10 cases where the targets were concealed out of the subject's reach, and one case (Haraldsson, 1978) where they were presumably stored only within an ESP test machine.

If clairvoyance is not much affected by distance, as is usually assumed, then there would seem little reason for placing the target in close proximity to the subject. This obviously increases the possibilities for trickery (as by substitution of envelopes). On the other hand, it does not usually permit any inadvertent leakage, because the subject is not generally allowed to handle the target container, or look behind screens. In cases where subjects were allowed to handle target envelopes, the envelopes were usually either sealed (Braud & Mellen, 1979; Parker & Beloff, 1970, Exp. 1; Schmeidler, 1971, Pilot) or, in one case, doubly enclosed (Braud & Braud, 1977).

There were two cases where handling of the target materials may have allowed sensory leakage (Johnson & Kanthamani, 1967, Exp. 1; Shields, 1962, Exp. 2). This was obviously true in the case of Shields' (1962) second experiment; the

subjects (school children) were allowed to directly touch the target cards, which were under a cloth framework. It is difficult to see how this can be considered an experiment in "extrasensory" perception, when the tactile sense was not excluded. In the experiment by Johnson and Kanthamani (1967, Exp. 1) the subjects were allowed not only to handle the unsealed envelopes, but to open them as well, after having made their calls. It is quite possible that subjects who worked rapidly would inadvertently begin opening an envelope before having completed the call. The experimenter, busy with her recording tasks, would not necessarily notice this departure from normal. This study, and the study by Shields, would both appear to be flawed, for the reasons outlined.

Usually, the targets were securely enclosed in "opaque" envelopes. In most studies the opaqueness of the envelopes was guaranteed by the construction of the materials (e.g., Braud & Mellen, 1979; Schmeidler, 1971, Pilot; Schmitt & Stanford, 1978) or by tests done to assess opacity (Honorton, 1972; Honorton & Stump, 1969; Parker & Beloff, 1970, Exp. 1). In other cases the envelopes were described as "opaque" with no further explanation (e.g., Carpenter, 1971, Exp. 2; Johnson & Kanthamani, 1967, Exp. 1; Kanthamani & Rao, 1971; Rao, Dukhan, & Rao, 1978). With some misgivings, I decided to assume the opacity of the envelopes, and not assign flaws in those cases. A further weakness in the studies by Kanthamani and Rao (1971) and Rao, Dukhan, and Rao (1978) was the enclosure of the key cards while the subject was present in the testing room. In the latter study the enclosure was done at a high table "completely out of the visual range of the subject" (p. 10). In the former study the enclosure was done underneath the testing table, also outside the subject's view. These procedures are secure, if one can assume that no tricksters were present in the subject pools. The sample studies were usually designed under the assumption that no such subjects would be present (see Section 5.2).

In a few studies the targets were not enclosed, but were merely screened from the subject (Fahler & Cadoret, 1958, Exp. C; Nicol & Humphrey, 1953; Sargent, 1978a; Shields, 1962, Exp. 1). These all involved variations on the "BT test," as described by Rhine and Pratt (1962, pp. 146-147). In this test the experimenter holds the cards behind a screen, and removes them one at a time as the subject makes the calls. The safest procedure is to remove a card from the top of the deck only after the subject makes the call; there can then be no possibility of reflections from the face of the card. In the study by Nicol and Humphrey (1953) the card was removed before the call. However, the cards were well-screened, so this would seem only a minor weakness.

There were more serious weaknesses in the studies by Fahler and Cadoret (1958, Exp. C) and Shields (1962, Exp. 1). In the Shields study, critical details were lacking, especially on the dimensions of the screen. The study must be considered flawed. In the study by Fahler and Cadoret, the cards were placed in a shallow box for screening purposes. They were then slid from one end of the box to the other, across a rounded middle partition, *as* the subject made the calls. Having examined a similar (possibly the same) apparatus once used in the Parapsychology Laboratory at Duke (where the experiment took place), I am

doubtful of whether sensory cues were excluded. This depends on the dimensions of the box, and on the relative locations of box and subject, but such information is lacking. With one subject the investigators used the safer "down through" technique, in which the deck of cards is left in an undisturbed pile as the subject makes the calls. Unfortunately, it was not stated where the shuffling and cutting of the card deck took place. Perhaps the subject was able to glimpse the bottom card, on occasion. At any rate, the study as a whole is hard to evaluate, given the reporting deficiencies. Hence I assigned a flaw.

There were similar reporting deficiencies in studies by Casler (1962), Nash and Nash (1961), and Wilson (1964, Exp. 1). In none of these studies was the relative location of subject and target(s) described. In the first two studies the "down through" technique was used exclusively, and the bottom card was shielded from view. I assumed that subjects were not permitted to handle the decks in either of these studies, and on that basis found the procedure secure. In the third study (Wilson, 1964, Exp. 1) the target was enclosed in an envelope which was apparently unsealed. Again, it was unclear whether the subjects were allowed to handle the envelope. Somewhat arbitrarily, I did assign a flaw in this study, since the procedure is less well-defined than the "down through" technique. Obviously, there are subjective judgments involved in assessing a "flaw" as opposed to a "minor contaminant" or a "procedural weakness." (It should be noted, in passing, that Wilson himself never claimed to have found evidence for ESP.)

In summary of the clairvoyance experiments, there were five cases (out of 33) which appear to have allowed possibilities for sensory leakage (Fahler & Cadoret, 1958, Exp. C; Johnson & Kanthamani, 1967, Exp. 1; Shields, 1962, Exps. 1 & 2; Wilson, 1964, Exp. 1). Three of these experiments had already been assigned flaws for their randomization procedures (Johnson & Kanthamani, 1967, Exp. 1; Shields, 1962, Exp. 1) or for a failure to describe randomization (Shields, 1962, Exp. 2). Hence the proportion of flawed experiments increases from 13/54 (in Section 3) to 15/54.

4.4. Sensory Leakage from Agent to Percipient

In telepathy experiments one must exclude the possibility that the subject learns about the target indirectly, through cues from the agent. For this reason Rhine and Pratt (1962) advised against telepathy testing: "If there is no special reason for including telepathy, . . . the test should by all means be carried out without a sender or agent present who is aware of the card order. Almost all of the difficulties that have arisen in controlling against sensory cues have come up in tests of GESP" (pp. 32-33).

Within my sample there were 20 experiments where all trials were with an agent, and another two experiments (Bevan, 1947; Shields, 1962, Exp. 1) where half the trials were with an agent. (Precognitive trials, which allow the tightest control over conditions, were used exclusively only in a study by Ryzl, 1968.)

In the discussion which follows, I will consider only inadvertent leakage from agent to percipient, leaving the question of deliberate cheating to Section 5.

In about half (12/22) of these experiments the percipient and agent were in separate rooms, with one of the rooms soundproofed in some manner. I have included a study by Krippner (1968) in this group, on the assumption that the Maimonides "sleep room," which was used in this study, was soundproofed. In five of the 12 cases rooms are reported as "soundproofed" but are otherwise undescribed (Braud & Braud, 1974, Exp. 2; Sargent, 1980b, Exps. 2, 3, & 5; Sargent, Bartlett, & Moss, 1982). In Sargent's experiments the agent moved at the end of the session from the soundproof room to a "quiet room" that was apparently unshielded. Possibly, the percipient had not at this time completed the judging. However, the "quiet room" was in a building which did not adjoin the percipient's building. Moreover, the agent in the "quiet room" was usually one of the experimenters. Hence there would seem to have been no possibility for inadvertent cueing. There was a change from these conditions in the fifth experiment from Sargent (1980b), where half the trials were conducted in unshielded rooms within the same building. Yet, results were marginally significant in the half conducted between two buildings. In all six of the cases cited above, an inadvertent cueing appears to have been eliminated, despite a vagueness in the experimental reports.

In six of the twelve experiments commercial sound-isolation rooms were used (Honorton & Harper, 1974; Moss & Gengerelli, 1968; Moss et al., 1970; Sondow, 1979; Terry & Honorton, 1976, Exps. 1 & 2). In four of these experiments this was not apparent from the reports (Honorton & Harper, 1974; Sondow, 1979; Terry & Honorton, 1976, Exps. 1 & 2). My information is based on a personal communication from Honorton (1981), and on a phone conversation with Sondow (1983). Auditory cueing was presumably eliminated, though this was not objectively assessed.

This leaves another 10 cases where percipient and agent were not in shielded rooms. In eight of these cases the percipient and agent were, at least, in nonadjoining rooms which were separated by distances ranging from 30 to 78 feet (where reported). These distances were probably sufficient, but I would feel more secure if tests for auditory transmission had been conducted. My concern is greatest in two experiments (Braud & Braud, 1973, Exp. 2; Braud & Wood, 1977) where the agents were someone other than the experimenters, and where the percipients made oral (as opposed to written or push-button) responses. The oral responses make the notion of a subtle guidance by the agent more plausible. In the Braud and Wood (1977) study, the percipient's verbalizations were actually fed to the agent through an intercom. However, the intercom operated only one-way. For the agent to have cued the percipient, any sounds would have had to travel a distance of 70 feet, through two closed doors, and would have also had to overcome the masking sounds of the "pink noise" which was fed through the percipient's earphones. In Braud and Braud (1973, Exp. 2) there was a distance of 78 feet separating agent and percipient. There was no intercom used in the latter study. In this study, as well as the last, it would appear that auditory

cues were eliminated. However, it would have been preferable if the experimenters had objectively assessed the extent to which such cues had been eliminated. In some buildings sounds travel quite readily between distant rooms.

Despite these reservations, I chose to assign a flaw only in studies by Casler (1964) and Shields (1962, Exp. 1), where there were obvious possibilities for sensory leakage. In Casler (1964) the agent and percipient were in adjoining rooms with both doors open, and the participants in visual contact with the experimenter (who remained in the hallway). Auditory cues were excluded by the dubious means of playing a phonograph record during the trials. In Shields (1962, Exp. 1) agent and percipient were in the same room directly across from one another. The cards were screened from the subject, but it is unclear whether the experimenter/agent's face was screened.

On a cumulative basis, this increases the proportion of flawed experiments to 16/54 (Shields, 1962, Exp. 1, having already been cited).

4.5. Sensory Leakage from Experimenter to Subject

Rosenthal and Rubin (1978) have reviewed many studies bearing on the question of whether experimenters can bias their subjects' responses in the direction of the experimental hypothesis. Presumably, this happens on an unintentional basis, for the most part, though the mechanism for the effects is not very well-understood. Commentaries following the Rosenthal and Rubin article indicate that there are disagreements as to the extent to which behavioral research has been contaminated by this artifact (see also Barber, 1976). Yet, there is no question but that these "experimenter expectancy effects" do occur, and have invalidated some experiments. For this reason, parapsychologists have generally taken the position that such bias should be ruled out by keeping the experimenter blind to the target order (e.g., Schmeidler, 1977a, p. 134).

Experimenter bias was an issue raised by Marks and Kammann (1980), in their critique of Targ and Puthoff's (1974) research with Uri Geller. Targ and Puthoff (1974) had described their communication with Geller as taking place through a "one-way audio monitor, operating only from the inside [of Geller's shielded room] to the outside" (p. 603). Since the experimenters could not communicate with Geller, experimenter bias was apparently eliminated. However, Marks and Kammann learned that the "monitor" was in fact a two-way intercom, by which Targ and Puthoff could communicate with Geller, by pushing a button. In a later report Puthoff and Targ (1976) describe two trials where "Geller was notified via intercom when the target picture was drawn and taped to the wall outside his enclosure" (p. 54). With this evidence Marks and Kammann labeled the description in Targ and Puthoff's (1974) earlier report as "not honest" (1980, p. 132). To my knowledge, the investigators have not publicly responded to that allegation. (See also Stokes, 1977.)

Hansel (1980) raised the issue of experimenter bias in a critique of a "dream telepathy" experiment by Ullman and Krippner (1970, pp. 98-104). Specifi-

cally, Hansel alleged that an experimenter may have been with the agent when the latter opened the target envelope for a night's session. If an experimenter did know the target, he or she could have inadvertently biased the subject's dream report in the direction of the target content.

Published accounts of the experiment by Ullman & Krippner (1970) and Ullman, Krippner, & Vaughan (1973) were ambiguous. However, I found a preliminary report of the experiment (Ullman & Krippner, 1969) which resolves the ambiguity. It is clear that the experimenter had no contact with the agent, after the agent had selected the target envelope. How then did the agent receive instructions for the night? (This was Hansel's question.) It turns out that the instructions were on a written form, enclosed in the envelope. If the published accounts had provided these facts, the controversy would never have arisen.

It appears, however, that the experiment can be criticized from another perspective. Generally, the experimenters were well aware of the need to keep everyone blind (e.g., Ullman & Krippner, 1970, pp. 67 & 83). However, on one trial from the experiment in question, there was a clear violation of this experimental requirement: An experimenter, possibly a research assistant, personally selected the last target after the agent had rejected two blind selections as "too gruesome" (Ullman, Krippner, & Vaughan, 1973, p. 151). Possibly, the experimenter did not see the target selected, but he at least saw which targets had been rejected. The trial should clearly have been excluded from the series, on the basis of nonrandom target selection, and nonblind conditions.

If there were no other violations of the experimental protocol, then Hansel's hypothesis (of leakage from the experimenter) can be rejected, since the results would remain significant with the questionable trial eliminated. Yet, the incident described leaves doubts as to the rigor with which the experiment was conducted.

4.6. Nonblind Experimenters in the Sample

Based on the 54-experiment sample it appears that most parapsychologists are well aware of the need to keep the experimenter blind to the targets. Two experiments by Shields (1962, Exp. 2) and Wilson (1964, Exp. 1) were the only ones that could be faulted on this basis. In both studies the experimenters who prepared the targets while observing their order, also administered the ESP test. (Neither Shields nor Wilson had published previous ESP experiments.)

There were three other studies where it was not entirely clear whether the experimenter was blind, but it seemed reasonable to suppose so (Honorton & Stump, 1969; Moss & Gengerelli, 1968; Moss et al., 1970). In the case of the Honorton and Stump study, I received assurances from Honorton that the experimenter was still blind to the target, at the time when judging took place.

There were two cases where the experimenter did not himself know the target, but had a brief encounter with someone who did know the target (Stanford & Mayer, 1974; Terry & Honorton, 1976, Exp. 2). (The situation in Terry and

Honorton's second experiment is more clearly described by Terry, 1975.) One could hypothesize a chain of nonverbal communication, and assign flaws. The encounters were so brief that I chose (with some misgivings) not to do so.

There were other remote possibilities for experimenter bias in the blind matching studies (Carpenter, 1971, Exp. 2; Kanthamani & Rao, 1971; Rao, Dukhan, & Rao, 1978). The reader may recall (Section 4.3) that in these studies the subject attempts to match target cards against key cards. The target cards are held face down, while the key cards are enclosed in envelopes. In a series of experiments by Kanthamani and Rao (1971) the experimenters shuffled the key cards under a table before inserting them into their envelopes. It seems unlikely that the experimenters could have formed any association between cards and envelopes, unless this was possible by tactile discrimination. In the other studies, visual association may have been possible, if the experimenters saw the key cards as they inserted them into their envelopes (this aspect of the procedure is not well-described). However, they could not have influenced the subjects' responses unless it is supposed that they (the experimenters) identified the target cards from their backs, as well as the key envelopes. This is unlikely. Experimenter bias is more or less ruled out in these studies, but it would have clearly been preferable if someone other than the testing experimenters had enclosed the key cards.

In summary, there were clear instances of nonblind experimenters only in experiments by Shields and Wilson. Since these experiments have already been cited for flaws in Sections 3.3 or 4.3, the proportion of flawed experiments remains at 16/54.

4.7. Sensory Leakage to Judges: "Remote Viewing" Experiments

In free-response ESP experiments sets of targets and responses (with the latter usually consisting of transcripts of the subjects' verbalizations, or copies of their drawings) are often submitted to blind judges for evaluation (by rating or ranking all possible target/transcript pairs). It then becomes critical to ensure that the judges have no information about the order in which the targets were presented. If they do, the judges may succeed in time-ordering both the target and the response sets, and match them on a non-ESP basis.

A controversy developed over the presence of such time-ordering cues in some of Targ and Puthoff's (1977) "remote viewing" experiments. The cues consisted of direct or indirect references to previous targets, which Marks and Kammann (1978, 1980) had found in some of the transcripts. For example, in the fifth transcript from a series with subject Pat Price, there was a reference to the fourth target (the Redwood City Marina). Marks and Kammann eliminated this and other references to previous targets, and supervised a rejudging. This time, judges were unable to match targets and transcripts with better than chance accuracy. However, the rejudging was not particularly sensitive, because it involved only five out of the nine transcripts.

Targ and Puthoff acknowledged the existence of the cues, but denied that this could account for the more striking correspondences that they had obtained. To prove their point, they had the transcripts edited with the help of Tart, and enlisted their own judge (who had no knowledge of the Price series). Using all nine transcripts and targets, the new judge was able to match seven out of the nine transcripts to their correct targets, a highly significant finding. Marks (1981b) objected to the procedures of the rejudging. He speculated that Tart might have been biased in his editing, or that the "blind" judge might have at some time encountered published reports of the Price series.

Readers who are interested in further details may consult the cited sources. The point to be made, however, is that no amount of editing can rectify the critical flaw in the procedure, which is the use of trial-by-trial feedback with a "closed" target pool. This problem was initially explained by Hyman (1977b):

> The statistics and judging procedure assume independence of descriptions for each target site. But this is obviously violated by the experimental procedure. Immediately after the subject generates his description, he is taken to the target site to be given feedback on how well he had done. Although the reason for doing this may be understandable, it makes his next description no longer independent of the first target site. To give one example of how this might generate false hits, assume that the first site is a municipal swimming pool. The next day the subject will probably avoid describing features that obviously belong to a swimming pool. If the second site, say, is a marina, the subject, in the third protocol, would avoid describing things that obviously belong to a swimming pool or a marina, and so on. Such a situation, in principle, could suffice to give a judge sufficient information to make perfect matches at each site from the descriptions . . . [p. 20].

Obviously, this is not a problem that can be solved by editing. Kennedy (1979b) has described two ways of avoiding the problem. The first is to use open-deck targets, and have judges evaluate transcripts in the order in which they were produced. The second solution is the one commonly applied, which is to have subjects do their own evaluations after each trial (again with open-deck targets).

Even though a wrong procedure was followed, the Price series may yet be salvaged. It should be possible to conduct a rejudging in which nine independent judges each evaluate one transcript against whichever targets the subject had not yet visited at the time of that transcript. This would be cumbersome but feasible.

4.8. Sensory Leakage to Judges: Handling Cues

Within my 54-experiment sample, there were no cases where judging procedures had been confounded by the use of trial-by-trial feedback. However, there were a number of cases where judges may have had handling cues available: Suppose (in an extreme case) that an agent handles the target and manages to spill coffee on it and also leave a deposit of cigarette ash. This target (perhaps an art print or photograph) is mixed with unhandled controls, and presented to the percipient who acts as a judge. He or she might easily identify the target on the

basis of these cues, despite having no ESP ability at all. In a more realistic example the cues would be something more subtle, such as a slight smudge or a crease on the target.

Hyman (1977a) saw the placing of such cues, which could be either inadvertent or deliberate, as a "clear-cut possibility" (p. 48) in a ganzfeld study by Honorton and Harper (1974). The targets in that study were Viewmaster reels, which the agent removed from the viewer, at the end of a session, to mix with control reels. Honorton (1979) addressed this question two years later, when the same issue was raised by Kennedy (1979a,b). (See also Stokes, 1978.) Honorton thought that the handling cue hypothesis was overly speculative. Even if such cues were present (and they were not known to be), they would not necessarily have been detected by subjects. To further support his case, Honorton did a comparison between studies that allowed handling cues, and studies that did not. He found no difference in their success rates (see Section 11.2, for a critique of this argument).

I believe that Honorton could more directly respond to Hyman and Kennedy by conducting a rejudging of this material (if the subjects' imagery reports are still available). Outside judges could compare the subjects' imagery with an unhandled set of Viewmaster reels, in the manner recommended by Kennedy (1979b). Until then, the study must be considered flawed, since the existence of such cues is more than just a remote possibility. If such cues did exist, they were not necessarily "subtle" or "subliminal," as assumed by Honorton; they may have been so blatant as to be readily detected by subjects. This is obviously the case if the cues were deliberately introduced (see Section 5.2). If the handling cue hypothesis is speculative, it is no more so than the ESP hypothesis.

The Honorton and Harper study was within my 54-experiment sample. Another such study was by Sondow (1979), where the handling cue hypothesis is also applicable. This possibility was noted by Sargent (1980a), an English parapsychologist. In reply to Sargent, Sondow (1980) did find holes in some (but not all) of his argument. She concluded that handling cues could not explain (a) the "dramatic quality of the correspondences between target and response" (p. 269), nor (b) the significant matching of targets and responses by outside judges. However, the outside judges were not always successful. In particular, they succeeded only when they had access to the subjects' "associations." The subjects' "associations" were obtained after they had viewed both the target and control pictures; hence these could have been influenced by handling cues.

The appeal to "dramatic correspondences" is one that was also made by Tart, Puthoff, and Targ (1980). I find it an uncomfortably subjective argument, particularly when it is based, as with Sondow (1980), on the publication of selected excerpts from subjects' reports. The assumption is that subjects have no knowledge of the target pool, but in Sondow's experiment such information could have been obtained from previous subjects. In any case, some "remarkable correspondences" can arise purely as a result of chance coincidence (Child & Levi, 1980). It is for this reason, of course, that parapsychologists typically rely on objective techniques of assessment.

4.9. Handling Cue Possibilities within the Sample

Altogether, there were 11 experiments within my sample where either the experimenter or an agent handled the target, and could conceivably have left cues that would differentiate target from controls (Braud & Braud, 1974, Exps. 1 & 2; Braud & Braud, 1977; Braud & Wood, 1977; Braud, Wood, & Braud, 1975; Honorton & Harper, 1974; Schmitt & Stanford, 1978; Sondow, 1979; Stanford & Mayer, 1974; Terry & Honorton, 1976, Exps. 1 & 2). In studies by Braud & Braud (1974) and Braud, Wood, and Braud (1975) the experimenters covered the target with a one-eighth-inch glass during sending, a procedure which eliminated some cues but may have introduced others (e.g., a flattening of the target). The procedure did not at all prevent handling cues before or after sending, and there is a further problem: The target envelope (unsealed and presumably resealed by the agent) may have differed from the control envelopes.

I also assigned flaws in two clairvoyance studies (Schmitt & Stanford, 1978; Stanford & Mayer, 1974), since the target was handled by an experimenter during target preparation, even though there was no agent.

I did not assign a flaw in a study by Braud and Wood (1977), since significant results were obtained with one technique (binary coding) which was completed before subjects viewed the target and controls. This leaves 10 experiments which should be considered flawed, until their results can be rejudged with duplicate target sets. Since only three of the studies have been cited for flaws in previous sections, the cumulative proportion of flawed experiments is increased from 16/54 to 23/54, in the sample as a whole.

4.10. Sensory Biases Affecting ESP Correlations

There is a type of sensory bias that can inflate correlations between ESP and other variables, even when the usual sources of sensory leakage have been eliminated. This can happen when subjects receive feedback on their ESP scores before being measured on the other variable. Consider, for example, a study by Palmer and Vassar (1974). The authors administered an ESP test, gave ESP feedback, and then administered a scale of mental imagery. Although a significant correlation was found between imaging ability and ESP, this finding was open to an alternative interpretation: Perhaps the subjects' knowledge of their ESP scores had influenced their imagery scores. In a further study, Palmer and Lieberman (1975) tested this interpretation by giving half of their subjects first an ESP test and then an imagery test, while reversing the test order for the remaining subjects. A significant correlation was found only in the former condition, and that correlation was significantly higher than the correlation obtained in the latter condition. The authors conclude as follows:

> These findings strongly suggest that the relationship between ESP and vividness of imagery found by Palmer and Vassar was artifactual, due either to a

> conscious effort on subjects' parts to confirm what they considered to be the experimenters' hypothesis, to differences in mood caused by learning of their success or failure on the ESP test, or to some other factor [p. 206].

Yet, in his review of ESP personality correlates Palmer (1977) cited a number of "significant" confirmations which were plagued by the same problem—the personality test having been given only after subjects obtained feedback on their ESP scores. Among the 12 studies of personality correlates in my sample, eight employed the questionable testing order (Casper, 1952; Johnson & Kanthamani, 1967, Exp. 1; Kanthamani & Rao, 1971, Pilot, Exps. A & B; Nash & Nash, 1961; Nicol & Humphrey, 1953; Shields, 1962, Exp. 1). All of these experiments must be considered flawed. It should be noted that personality test scores are more easily influenced by situational variables than is ordinarily supposed. For example, there is a general tendency for such scores to "improve" between test and retest, even though no experimental manipulation has been introduced (Windle, 1954). The reasons for the "improvement effect" are not well-understood.

In several of Schmeidler's group studies of the "sheep-goat effect," there may have been a similar contaminant. However, this possibility was clearly ruled out in at least one group session, where significant results were obtained (Schmeidler & McConnell, 1958, pp. 125-126). In the individual experiments by Schmeidler, subjects were always classified on attitude before receiving ESP feedback, but they were sometimes reclassified on the basis of remarks made after feedback (Schmeidler, 1959). Further discussion of this issue will be deferred to Section 7.2.

Of the eight experiments which permitted feedback artifacts, three had already been assigned flaws in previous sections (Casper, 1952; Johnson & Kanthamani, 1967, Exp. 1; Shields, 1962, Exp. 1). Hence the overall proportion of flawed experiments is increased from 23/54 to 28/54.

There is some interest in considering the combined effects of sensory leakage, while ignoring problems introduced by randomization failures. Out of the 54 experiments in "extrasensory perception," 41 percent (22/54) allowed some possibility of sensory influence on the results. The sources of this influence were primarily handling cues left on the target and cues from ESP feedback.

5. Subject Cheating

5.1. Introduction

Many claims for paranormal phenomena have been based on tests with highly selected subjects (e.g., psychics, sensitives, or mediums). Obviously, a first requirement of such tests is that they rule out trickery by the subjects. Hansel (1980), pp. 29-33) cites several early experiments where trickery was not ruled out, and where the subjects later confessed to having used tricks.

One of the most remarkable tricksters was Margery Crandon, a medium who

succeeded in duping prominent American researchers of the 1920s and 1930s (Hansel, 1980, pp. 75-78). Rhine was one researcher who was not duped; he investigated Margery and immediately detected trickery (Mauskopf & McVaugh, 1980, pp. 75-77). Yet, members of the American Society for Psychical Research (ASPR) were consistently fooled, and published exuberant accounts of Margery's psychic feats in their *Journal*, long after she had been exposed by Rhine and others. The Margery affair led to a split among psychical researchers, with the hardliners such as Rhine breaking away from the ASPR. It was years before this split was healed, with the ASPR coming under new leadership.

Given this historical context, one would expect modern researchers to be sensitive to the issue of subject fraud, and to carefully guard against such a possibility. Yet, there are indications that this is not always the case. In the 1970s, a number of leading parapsychologists (e.g., Beloff, 1975; Eisenbud, 1976) began endorsing Uri Geller as a genuine psychic. Geller's most publicized claim was of his ability to bend metal objects (e.g., a spoon or a key) through the use of paranormal powers. Many observations of these phenomena were made, and they were published in a collection edited by Panati (1976). Eisenbud (1976) judged that the papers in the collection were of "high caliber," leaving "little room for doubt" about Geller's metal-bending ability (p. 322). Yet, the papers were almost all based on uncontrolled observations, and had little more than anecdotal value. Weakness in the papers were exposed by Hyman (1976), who wrote a devastating critique of the Panati collection. Hyman found that investigators of Geller had, generally speaking, made little effort to control against trickery. Control against trickery was critical, since Randi, a professional magician, had succeeded in duplicating most of the phenomena (Randi, 1975). As it became obvious that trickery had not been excluded, hardline parapsychologists began ignoring the Geller phenomena. Thus, Stanford (1977b), in his Wolman *Handbook* chapter on psychokinesis, only briefly refers to this research, which he sees as "still in need of tightening up" (p. 329).

Geller was not the last of the presumed metal-benders. At the 1981 Convention of the Parapsychological Association, there was a report titled, "Exploratory Research with Two New Psychic Metal-Benders" (Phillips & Shafer, 1982). The report was from the McDonnell Laboratory for Psychical Research, which is a relatively new center for parapsychological investigation. The authors strongly implied that the phenomena were genuine, though they did not claim conclusive evidence. It was well that they did not. Magician Randi later revealed that it was all a hoax; he had planted the "metal-benders" in the McDonnell lab (Broad, 1983). These "psychics" were merely clever tricksters. Phillips, the director of the McDonnell lab, admitted that he had initially been fooled. He pointed out, however, that his research team had never issued a formal report, since it had been unable to elicit the phenomena when controls against trickery were tightened (Broad, 1983). The need for such controls is evident from the limited success that Randi's hoax achieved.

5.2. Subject Cheating in the Sample

In my 54-experiment sample the subjects were unselected volunteers, rather than highly touted "psychics." On that basis cheating should have been less of a problem. A study by Moss and Gengerelli (1968) did include five professional psychics among the 144 subjects, but the five did not score especially high. There were other studies where the subjects were semi-selected, in that they were chosen for having been successful in previous ESP experiments (e.g., Moss et al., 1970; Sargent, 1980b, Exps. 3 & 5). However, the subjects in these studies were not offered any incentives for high scores. Such incentives were offered in studies by Haraldsson (1978) and Wilson (1964), where fraud is a more serious possibility. Fraud is also a concern in a study by Casler (1964) where the agent was paid by funds from the Duke Parapsychology Laboratory (amount unknown). Of course, there are many motives for fraud besides money.

In a number of the studies the volunteers had a special interest in parapsychology; they were members of parapsychology classes or societies (e.g., Braud, Wood, & Braud, 1975; Sargent, 1978a; Terry & Honorton, 1976, Exp. 1). I do not believe that such groups would normally include tricksters. However, the subjects would probably be more highly motivated than those from an unselected sample. This in itself might increase the risk of fraud.

In assessing the likelihood of subject fraud, I was less concerned with motives for fraud, than with the question of whether fraud was excluded. A motive for fraud is always present for some subjects, who get a kick out of fooling their experimenters (especially when the latter have claimed fraud-proof conditions). Estabrooks (1947, pp. 122-126) recalled how two college sophomores succeeded in fooling him, even though he knew that cheating would be attempted. This incident, discussed by Nicol (1976), occurred under formal testing conditions in Estabrooks' Harvard laboratory. The conditions were actually quite similar to those in some studies from the sample. In one such test the agent and percipient were 100 feet apart, with four closed doors separating their rooms. Despite close supervision by Estabrooks the agent succeeded in passing a signal, utilizing the light which shone through a crack under a door. A confederate, usually hidden in another room, assisted in conveying the information to the percipient, this time by way of an auditory code. The system was so successful that the percipient was able to guess the correct color of a deck of playing cards, on all 52 attempts.

It is true that, among the sample telepathy experiments, shielded rooms were used in 12 cases (see Section 4.5). In a few cases the rooms were reported as electrically shielded (Braud & Braud, 1974, Exp. 2; Terry & Honorton, 1976, Exps. 1 & 2; Sondow, 1979). This is important if subjects make use of miniaturized transmitters and receivers (which are routinely used by some students for classroom cheating). On the other hand, shielding is a useful control only when its effectiveness is objectively assessed at the time of the experiment. Often, shielding characteristics change over time, as modifications are made to experimental rooms (e.g., for insertion of cables or ventilation ducts). Sondow has in-

formed me that the room used in her 1979 study had been modified. Of course, a trickster could introduce his or her own modifications.

It should be mentioned that possibilities for cheating were considerably lessened, in seven telepathy experiments, by using an experimenter or an experimental assistant as the agent (Bevan, 1947; Braud & Braud, 1974, Exps. 1 & 2; Braud, Wood, & Braud, 1975; Grela, 1945; Krippner, 1968; Shields, 1962, Exp. 1). There were six other experiments where an experimenter served as agent for at least some trials (Honorton & Harper, 1974; Sargent, 1980b, Exps. 2, 3, & 5; Sargent, Bartlett, & Moss, 1982; Sondow, 1979). The results in two of these experiments (Honorton & Harper, 1974; Sondow, 1979) remain significant when one excludes trials where a friend of the subject served as agent. (See Rogo, 1979a, with respect to Honorton & Harper, 1974.) In the remaining experiments, authored by Sargent, the results are not broken down according to who the agent was.

In all of the above experiments, scenarios for fraud could be imagined, particularly where the subject was left unattended as in studies by Bevan (1947), Braud and Braud (1974), Braud, Wood, and Braud (1975), and Grela (1945). However, this assumes that tricksters were present in the samples. In my own experience, I have rarely encountered sophisticated trickery, even among subjects claiming psychic skills. On that basis I did not assign flaws. Braud and Braud (1974) show some awareness of the problem, since they locked the subjects into their rooms during the experiment.

Trickery may also have been possible in many of the clairvoyance experiments, though it is difficult to judge. A trickster might, for example, prepare duplicate envelopes, similar to those used to enclose the target. As the experiment was about to begin, he or she would find a means of misdirecting the experimenter's attention, and substitute the duplicate envelope for the experimenter's. The trickster would then need to open the target envelope and reseal it in a new container, all without the experimenter's noticing. Yet, an experienced magician could probably accomplish this. The method would fail if the target envelope had been sealed or coded in such a manner that the substitution would be detected.

I do not know whether the trick described, or a variety of other tricks, could have been used in the sample experiments. But in general, the sample experiments seem to have been designed under the (reasonable) assumption that a trickster would not be present. Usually, this would be a safe assumption. Hence, in assigning flaws, I was only concerned with cases where ordinary subjects might have cheated spontaneously, without much forethought. On that basis, I found eight experiments where cheating may have been an easy task (Braud & Braud, 1973, Exp. 2; Braud & Braud, 1977; Casler, 1964; Casper, 1952; Johnson & Kanthamani, 1967, Exp. 1; Schmeidler, 1970; Terry & Honorton, 1976, Exps. 1 & 2).

In the study by Braud and Braud (1973, Exp. 2), which was less controlled than their later research, the agents were allowed to select their own targets. Apparently, they were unsupervised, and could have made selections which

were not random, but based on the percipients' preferences. (The percipients were their spouses.)

In one of the telepathy experiments (Casler, 1964) the agent and percipient were at close range (see Section 4.4) and could easily have exchanged auditory signals. In three other telepathy experiments (Casper, 1952; Terry & Honorton, 1976, Exps. 1 & 2) agent and percipient were in separate rooms, but one or the other was unsupervised. In the study by Casper (1952) the agent could have easily signaled the percipient by varying the manner of operating a "ready" light. Where such "ready" signals are used, they should operate only from the percipient's room to the agent's room, so that cueing is eliminated. There were also obvious possibilities for cheating in the first experiment by Terry and Honorton (1976), where undergraduate subjects took turns testing each other. The senior investigators did supervise target selection, in some trials, but they apparently had no further role in controlling percipients, agents, and student experimenters. In the second study by Terry and Honorton, Terry was the sole experimenter (see Terry, 1975). He could not have simultaneously controlled both agent and percipient. In both of the Terry and Honorton studies, there was another possibility—that an agent deliberately introduced handling cues on the target (see Section 4.9).

Cheating was also a clear possibility in three clairvoyance experiments (Braud & Braud, 1977; Johnson & Kanthamani, 1967, Exp. 1; Schmeidler, 1970). In the study by Braud and Braud (1977) there was again the possibility that subjects would deliberately introduce handling cues. The authors advanced five arguments against that hypothesis, but none of these is entirely convincing. The possibilities for cheating were more direct in the study by Johnson and Kanthamani (1967, Exp. 1), where the subjects could have peeked into the unsealed envelopes while the experimenter was occupied with her recording task (see Section 4.3). There was also a direct avenue for cheating in an experiment by Schmeidler (1970), where subjects recorded the targets before having submitted their guesses. Perhaps they made small "revisions" in their guesses after learning the targets. The "revisions" might correspond with a subject's "second guess"; they would not have to be outright fabrications.

In addition to the eight experiments cited above, there were another four cases where it was difficult to judge whether subject fraud could have been easily accomplished (Haraldsson, 1978; Fahler & Cadoret, 1958, Exp. C; Shields, 1962, Exp. 2; Wilson, 1964, Exp. 1). In the Shields (1962, Exp. 2) and Fahler and Cadoret (1958, Exp. C) studies, the experimental apparatus is only vaguely described; it is unclear whether "peeking" was excluded. Likewise, it is unclear from Wilson's (1964, Exp. 1) report whether the subjects may have had an opportunity to open the target envelope (which was apparently unsealed). Haraldsson's (1978) test machine was undescribed, so there is no way of knowing whether subjects may have reset counters, or employed some other trick (a long-playing record was awarded for high scores). All four experiments may have failed to exclude cheating.

In summary, there were 12 experiments where subject fraud was not at all

excluded (or where this judgment could not be made). Since 11 of the experiments had already been cited for other flaws, the cumulative proportion of flawed experiments is increased only from 28/54 to 29/54 (by the inclusion of Schmeidler, 1970).

6. Recording Errors

6.1. Blind Recording in ESP Experiments

J.L. Kennedy, who had earlier documented the existence of cues on the backs of ESP cards (see Section 4.2), became involved in studies of recording bias. In a series of experiments he showed that recorders of ESP cards tended to make errors in the direction of their expectations (Uphoff & Kennedy, 1939). Later research has confirmed the existence of such biases (Barber, 1976; Rosenthal, 1978). The effects of the bias are generally small, but this depends on who the recorder is, and on how difficult the task is. In most of the "classic" ESP tests, the deviations from chance were too large to be accounted for in this manner (Murphy, 1938). (See, however, Scott, 1980, for an apparent exception.)

There were some ESP studies which could be so explained, and Kennedy persisted in his criticism. In response, parapsychologists began keeping the experimenter blind to the subjects' calls while recording the targets, and blind to the targets while recording the calls. Schmeidler (1977a) sees these blind recording techniques as an essential requirement of an ESP experiment, so long as recording has not been automated.

6.2. Nonblind Recording in the Sample

Among the 54 experiments in my sample, recording was clearly non-blind in seven cases (Casler, 1964; Johnson & Kanthamani, 1967, Exp. 1; Kanthamani & Rao, 1971, Pilot, Exps. A, B, & C; Shields, 1962, Exp. 2). In another seven cases recording was not well-described, and may have been nonblind (Carpenter, 1971; Casler, 1962; Haraldsson, 1978; Rao, Dukhan, & Rao, 1978, Pilot, Exps. 1 & 2; Shields, 1962, Exp. 1).

Most of the above experiments would seem to be flawed, but there are these exceptions: In Casler's (1964) experiment the targets were recorded on two separate occasions, which allowed a check on recording error. In the Kanthamani and Rao (1971) study the target cards were read aloud one by one as they were recorded (see Kanthamani, 1969, pp. 72-74). Rao informs me that the record sheets were later checked, to make sure that all target symbols appeared with equal frequencies (as would be expected with a closed-deck sequence). On that basis, I have not assigned a flaw, though the procedure used was a weak one.

Several of the cited experiments were blind-matching tests, where the subjects

placed cards in piles under key cards, according to where they thought a match might occur. In such an experiment it is important to carefully segregate the piles before exposing the key cards. Otherwise, there may be some ambiguously placed cards whose status as hits or misses must be decided on a post hoc basis. The result could be an artifactual psi-hitting or psi-missing effect, depending on the direction of the experimenter's bias. This is a possibility in experiments by Carpenter (1971) and Rao, Dukhan, and Rao (1978, Pilot, Exps. 1 & 2). Rao has informed me that he observed some of the initial sessions conducted by Dukhan (his student). If it can be established that these witnessed results were independently significant, then this counterhypothesis will be eliminated. My concern is partially derived from the fact that Dukhan collected an enormous amount of data over a surprisingly short time span (see Dukhan & Rao, 1973). There may have been a temptation to take shortcuts in recording, as well as in other aspects of the procedure (which are not described).

In a blind-matching experiment by Shields (1962, Exp. 2) it is unclear whether the individual target cards were recorded at all. Possibly, the experimenter simply made a visual search for matches, and recorded only the total hits from each run. Obviously, this procedure would enhance the possibilities for errors. Such errors may also have inflated results in a study by Casler (1962), where there is a similar failure to describe recording. A pertinent fact is that these experiments were the first published by Casler and Shields; the two investigators were relatively inexperienced.

Haraldsson (1978) used student experimenters who would also have been inexperienced. They used an ESP test machine, but probably had to hand-copy readings from digital counters on the machine (which was not described). It may be that subjects were allowed access to reset buttons. When this is allowed, some subjects will reset to zero before the experimenter has had a chance to make careful observations of the counters. Suppose, however, that recording was completely automated (as advocated by Gardner, 1975, and Hansel, 1980). It is still the case that no control data were collected during the experiment (or at least none were reported). There is no assurance that the recording equipment functioned as it was supposed to function.

In summary, there are nine experiments which may have been flawed on the basis of recording procedures: Carpenter (1971, Exp. 2); Casler (1962); Haraldsson (1978); Johnson and Kanthamani (1967, Exp. 1); Rao, Dukhan, and Rao (1978, Pilot, Exps. 1 & 2); and Shields (1962, Exps. 1 & 2).

There was one borderline case, a card experiment by Sargent (1978). Sargent conducted a clairvoyance test in which he removed cards one at a time from a pile *after* the subject made the call. This is the usual "BT technique" (Rhine & Pratt, 1962), except that Sargent recorded each card as he proceeded through the deck. The usual procedure is to record the deck of cards only after each run is completed. There are problems that can arise with Sargent's modified BT procedure: Trying this myself, I found that it was easy to get out of step with a subject who made rapid calls (the trials were not timed). Sometimes I would turn a card over *before* the subject made his call. When this happened I had

changed the conditions from clairvoyance to telepathy, which made cueing more likely (though Sargent's subjects were screened from both cards and experimenter). More important, I was no longer blind to the target while recording the subject's call. Assuming that Sargent was more practiced than I, in this technique, he may not have encountered the same difficulty. Nonetheless, I have chosen to assign a flaw here. There was no fixed intertrial interval in Sargent's experiment, and in such cases it is much preferred that recording take place at the end of the run, as is usual.

Of the 10 experiments discussed above, four had already been cited for flaws in previous sections. Hence the overall proportion of flawed experiments is increased from 29/54 to 35/54.

6.3. Automated Recording in ESP and PK Experiments

Within my sample of ESP experiments, there were only two cases of automated recording (Haraldsson, 1978; McBain et al., 1970). This is explained by two features of the sample. First, the sample included 20 studies published prior to 1970, when automated recording was rare. These were mostly card-guessing studies. Second, the sample included many recent free-response studies; of the post-1970 studies 20 out of 34 fall into that category. In free-response experiments, only a few trials are collected from each subject (often only one), and there are rather lengthy intertrial intervals. Under such conditions there is little need for automated recording devices.

The situation is different in PK research, where the targets must be obtained by some random process whose outcome cannot be known. Almost always this involves the use of an electronic REG (see Sections 3.4 and 3.6). In early research with such devices recording was not always fully automated; it was often necessary for the experimenter to make repeated readings of digital counters, and hand-copy these. Gardner (1975, 1981) has sharply criticized this procedure, since it allows obvious possibilities for human bias. In recent years the REGs have been increasingly used in combination with computers, so that hand recording of any kind is less common. (It should be noted that automated equipment does not entirely eliminate observational error, since this can enter at a later stage, as in the reading of a computer printout; see Rosenthal, 1976.)

The use of automated recording devices represents a major advance over earlier PK dice studies, where observational error was a major counterhypothesis (Girden, 1962). However, the limitations of such devices are not always recognized. Advocates of automated testing within parapsychology (e.g., Davis, 1974; Broughton, 1982) advise against a blind reliance on equipment. From the critic's side, Hansel (1980) concludes that "automation in itself is of little value without adequate experimental design" (p. 234). It is critical to good experimental design that automated equipment be subject to periodic checks, during the course of an experiment, to ensure that the equipment functions as it was intended to function. Yet, a 1980 survey of PK experiments by May, Humphrey,

and Hubbard (see Section 3.6) indicates that parapsychologists have generally not provided for such checks. There are, however, signs that the situation is improving, with some experimenters introducing control trials, generated at the same time as experimental trials (e.g., Broughton, 1979, 1982).

7. Classification and Scoring Errors

7.1. Hansel's Critique of Schmidt

Some early ESP experiments by Schmidt (1969a,b) are often cited as examples of good automated designs. Recording in these experiments was secure, since scores were kept on nonresettable counters within the test machine (which was periodically checked), as well as on a paper-tape output. However, Hansel (1980) claims to have found various possibilities for experimental error, particularly in the assignment of scores to high-aim and low-aim categories. It is not always clear, in Hansel's critique, where he is suggesting experimenter bias in the assignment, and where he is suggesting experimenter fraud. Restricting myself to the former question (i.e., the possibility of careless error), I have inferred three major reasons for Hansel's concern: (a) the internal counters, though nonresettable, could be switched out of circuit so that "practice" trials would not be recorded; (b) the paper tape was not continuous, but in disjoint pieces, some of which may have been left out of the final analysis; and (c) high-aim and low-aim runs were recorded on the same set of counters; the two types of scores canceled each other out so that their individual contributions could not be discerned.

The first criticism does not seem relevant, since it is perfectly legitimate to disconnect the counters during practice; this is necessarily a decision made before Schmidt knew whether the session would yield hitting or missing. Hansel's second and third points do seem relevant, though they do not appear to represent "critical flaws" in the procedure. It is true that the paper tape was not in a continuous strip (and the pieces had not been prenumbered). However, contrary to Hansel's implication, the number of hits on the counters was checked against the numbers on the paper tape, and these were found to be in agreement (e.g., Schmidt, 1969b, p. 103). It is not entirely clear whether these checks were made throughout an experiment, or only on some occasions. If it was the latter, then Hansel's criticism may perhaps be applicable. With regard to the third criticism, Hansel is correct, in that the counter readings did not distinguish between high- and low-aim runs; it would have been preferable if separate sets of counters had been employed for the two types of trials. Nonetheless, high- and low-aim trials were distinguished on the paper tape, as Hansel acknowledges. This was done by the setting of a switch prior to each run; hence it was a decision made in advance, and could not have been influenced by the success or failure of the run.

For the reasons cited above, I believe that Schmidt's early experiments do

deserve the attention they have received. However, I have one reservation: As Hansel emphasizes, Schmidt worked largely alone. The possibility of human error could have been lessened if additional experimenters had participated. Schmidt (1980) has himself suggested a design for the exclusion of such errors (including even deliberate fabrication of data), and has recently taken a step toward implementing this design (Schmidt, Morris, & Rudolph, 1983).

7.2. Scott's Critique of Schmeidler and McConnell (1958)

Classification error has long been an issue in parapsychological research. In the 1940s, researchers began classifying subjects into "psi-hitters" and "psi-missers," even when overall scoring was at chance. The evidence for an ESP or PK effect then depended entirely on the validity of the classification. In particular, a classification based in any way on *post hoc* knowledge of subject's scores could be worthless. The classification has to be blind and would ordinarily be completed before the ESP tests were carried out. Hansel (1980, p. 202) has made a valuable suggestion–that the basis for the division into high and low scorers be made public before any testing begins. There could then be no question as to whether subjects had been illegitimately reclassified, after ESP scores became known.

There have been many parapsychological studies where subjects have been classified as "sheep" (who are expected to score high) and "goats" (who are expected to score low). In the original research (Schmeidler & McConnell, 1958) "sheep" were those who believed in ESP (or were at least neutral towards the possibility), and "goats" were the disbelievers. Many comparisons between sheep and goats were made, and the sheep almost always obtained a slightly higher average than the goats. Overall, the difference was highly significant.

In a review of the Schmeidler and McConnell volume, Scott (1959) suggested two counterhypotheses: (a) that the data of some subjects had been illegitimately excluded, and (b) that the experimenter (Schmeidler) had been biased by knowledge of the ESP scores in deciding when to reclassify subjects (from sheep to goats or goats to sheep). In answer to Scott's first objection, Schmeidler (1959) published a list of all records which had been omitted from the original analyses. Assuming that her records were well kept over the years, Schmeidler's defense may have succeeded here. However, the situation is less clear with respect to Scott's second objection. Schmeidler (1959) asserted that in all the individually administered tests, ambiguities in the subject's status (as sheep or goat) had been resolved prior to the ESP test. Yet, Schmeidler's list of omitted subjects (Table 1) indicates that there were some exceptions to this rule. Were reclassified subjects always left out of the analysis? Scott notes that in one early experiment (Schmeidler, 1943) there were four subjects whose results were reclassified and then left in the analysis.

Early accounts of the individual experiments (Schmeidler, 1943, 1945; Schmeidler & Murphy, 1946) do not describe the procedure in ambiguous cases

(where subjects could not be readily classified as either sheep or goats). Ambiguous cases may have been common, since the criteria for classification were loose. Schmeidler and Murphy (1946) described sheep as those who accept the theoretical possibility of ESP. In another report of the same experiments (Schmeidler, 1945) sheep are described as those who accept that ESP could occur in the actual experimental situation. However, Table 1 of the latter source described sheep as those who hope to succeed in the experiment. Schmeidler and Murphy (1946) wrote that "a rigid phrasing of the question was not considered necessary" (p. 271). There were no fixed response alternatives, nor a preplanned scoring system.

In Schmeidler's group experiments the assignment of subjects to the sheep or goat categories depended on integrating several sources of information. This varied from one experiment to the next. In all experiments the major datum was the subject's self-assessment as a sheep or a goat, but this might change during the testing session. In some experiments, subjects were asked to give an open-ended explanation of why they considered themselves sheep or goats. In other experiments they were asked to mark their degree of belief on a line representing a continuum from belief to disbelief. Precisely how these sources were integrated, from one experiment to the next, is unclear from the published reports. McConnell (1959) claims that the sheep-goat forms from a September 1949 series were such that no ambiguities arose. Unfortunately, that particular series may have been one where subjects obtained ESP feedback before having submitted the sheep-goat forms. They may have made last-minute changes in their self-assignment, after having learned their scores (see Schmeidler & McConnell, 1958, Appendix B).

In many of the group series, the ESP data were recorded on sheets which were separate from the sheep-goat forms. On this basis, Schmeidler (1959) rejects the possibility of *post hoc* assignment. However, all records had identifying names or numbers, and it would not be surprising if the experimenter associated certain names or numbers with unusually high or low ESP scores. Where Schmeidler definitely knew the ESP scores, she took the precaution of submitting the sheep-goat form to a third party, such as Murphy, for evaluation (Schmeidler & McConnell, 1958, pp. 48, 113). However, Scott notes that the decision to submit or not submit an ambiguous form could itself have been inadvertently biased by knowledge of the ESP score.

7.3. Classification and Scoring Errors in the Sample

The Schmeidler experiments, which were part of my sample, were considered flawed on the basis of Scott's critique. There were three other experiments in the sample where the scoring of sheep-goat data, and the assignment of ambiguous cases, was again unclear (Carpenter, 1971, Exp. 2; Moss & Gengerelli, 1968; Schmeidler, 1971, Pilot). In Schmeidler's (1971) study, the sheep-goat responses were scored first by the experimenter (Lowy), and apparently rescored

by Schmeidler, using Lowy's notes. It is unclear from the report whether Lowy or Schmeidler had seen the subjects' ESP scores, prior to classification. A similar question could be raised with respect to Bhadra's (1966) experiment, in which a division between sheep and goats was accomplished only after data collection. However, there were, at least, fixed response alternatives in Bhadra's study, and the scoring system devised appears to be a reasonable one. Hence I chose not to assign a flaw in this borderline case.

My sample also included two studies of personality correlates where the classification into criterion groups was nonblind (Johnson & Kanthamani, 1967, Exp. 1; Shields, 1962, Exp. 1). In Shields' experiment, the division into "withdrawn" and "not-withdrawn" groups was based in part on projective personality tests which were administered and scored after the experimenter knew the ESP scores. There was a similar problem in a study by Johnson and Kanthamani (1967, Exp. 1), and this was acknowledged by the authors. In a second experiment, they took precautions to ensure that scoring of the projective measure was blind. An .05-level effect emerged in the same direction, but it did not satisfy Palmer's (1977) two-tailed criterion, and hence that study was not included in my sample (see Section 2).

A study by Sargent and Harley (1981) represents a borderline case. The authors conducted a pilot and two confirmatory experiments. They then conducted some "preplanned" analyses of the combined data from all three experiments. Those subjects who were high on extraversion (scores of 7 or 8) obtained higher ESP scores than subjects who were low on extraversion (scores of 0, 1, or 2). Subjects with intermediate scores were left out of the analysis.

The critical question is whether the cut-off points were selected prior to any collection of data. The analysis was "preplanned," but how far in advance of the analysis? The planning may have occurred before the pilot experiment, but it may also have taken place before the confirmatory experiments, or before the combining of data from all three experiments. Only if the planning was before the pilot, was bias in the choice of cut-off points eliminated. Yet, the pilot was described simply as a "preliminary examination of the value of the questionnaire" where the authors "did not expect any significant effects" (p. 202). For this reason, I am inclined to assume that the planning of the overview analysis took place at a later time, and that bias was not eliminated. Hence the study must tentatively be considered flawed.

In contrast with the personality measures, the ESP variables could usually be scored in a straightforward fashion. Dale (1943) found that the most common error, in fixed-choice experiments, is a failure to circle hits, resulting in a psi-missing effect. I suspected such errors in two experiments which showed overall psi-missing (Kanthamani & Rao, 1971, Exp. C; Ryzl, 1968). At my request, Rao supervised a rescoring of Kanthamani & Rao's data. Although a few such errors were found, these did not account for the results. Until Ryzl's (1968) data can be rechecked, I will be inclined to suspect such errors. It was necessary to hand-score 40,000 trials, and the scorer quite possibly knew which group (sheep or goats) a given record sheet belonged to.

In summary of results from the sample, there were nine experiments with possibilities for classification or scoring errors (Carpenter, 1971, Exp. 2; Johnson & Kanthamani, 1967, Exp. 1; Moss & Gengerelli, 1968; Ryzl, 1968; Sargent & Harley, 1981; Schmeidler, 1971; Schmeidler & McConnell, 1958, Individual & Group Exps.; Shields, 1962, Exp. 1). Four of the experiments had already been cited for flaws in previous sections. Hence the overall proportion of flawed experiments is increased from 35/54 to 40/54.

8. Statistical Violations

8.1. Criticism of the Statistical Model

The initial response of psychologists to Rhine's 1934 monograph focused on statistical issues (e.g., Kellogg, 1937). For the most part, these issues were resolved in Rhine's favor (Mauskopf & McVaugh, 1979); his results could not be dismissed as statistical artifacts. Critics of Rhine then began attending to issues of experimental design, where he was more vulnerable.

In the 1950s there was a revival of statistical criticism (e.g., Boring, 1955; Bridgman, 1956; Brown, 1957; Girden, 1962). Although their views differed considerably, all these authors denied that parapsychological claims could be proved by comparing ESP or PK test results with a theoretical probability distribution. The most developed argument was Brown's. However, it appears that Scott (1958) successfully refuted most, if not all, of Brown's argument (some of which involves mathematics that are beyond my understanding).

Later, the probability model was criticized by Gatlin (1977, 1979) and by Hardy, Harvie, and Koestler (1973). These works will not be reviewed here, since they are not so much methodological criticisms, as alternative interpretations of psi phenomena. Gatlin argues that much ESP data can be explained as a sophisticated form of pattern recognition. Hardy et al. (1973) see ESP as a form of meaningful coincidence. (See Kelly, 1974, for a critique of Hardy et al.; Gatlin's thesis has been critiqued by Pratt, 1979a, and Tart, 1979b.)

Several other recent critiques of parapsychology's statistical methods do include methodological recommendations (Alcock, 1981; Calkins, 1980; Girden, 1978; Moss & Butler, 1978). These authors have all argued for some kind of empirical control condition, such as a condition in which subjects do not wish for a particular outcome. Evidence for psi would be obtained only when scores in the experimental condition were significantly higher than those in the control condition. Alcock agrees with Boring (1955) that it might be impossible to "turn off" the hypothetical psi ability, in a control condition, but he nevertheless sees the inclusion of such a condition as necessary.

In response to Alcock, Morris (1982) observed that many "process-oriented" studies in the parapsychological journals do include control groups. In most such studies, the researchers do not claim to have "turned off" the psi variable,

but only to have lessened its chance for expression, in the control condition. In hypnosis research, for example, a control group might include subjects who had not been hypnotized. When I examined my sample studies, I found support for Morris' argument; in about two-thirds of the cases, the significance of the results did depend on comparisons between two experimental conditions or between two subject populations who were expected to score differently. Both Morris (1982) and Palmer (1981) see the inclusion of control or comparison groups as desirable in process-oriented research. Alcock sees their inclusion as an absolute necessity. From a practical standpoint this is not such a large difference in opinion.

Based on Palmer's (1981) arguments, I have not assessed a flaw in sample studies which lacked an empirical control group. As Palmer observes, the inclusion of a control group does not in any way reduce one's reliance on probability theory. Tests of the difference between experimental and control groups are necessarily grounded in probability assumptions. These assumptions were often checked in early ESP research, by comparing calls against targets for which they were not intended (Pratt et al., 1940/1966). The results nearly always conformed to what one would expect from probability theory.

If probability theory is flawed, and psi effects are some kind of statistical fluke, it then becomes difficult to account for the "experimenter effect" (Kennedy & Taddonio, 1976; Parker, 1978; White, 1977). This "effect" is simply a dependence of psi results on the identity of the person conducting (or supervising) the research. Many researchers, such as myself, obtain almost uniformly chance results, while others obtain extrachance results on a routine basis (e.g., Sargent, in his ganzfeld research). So far as I can tell, the successful and unsuccessful experimenters employ the same statistical techniques, as described by Burdick and Kelly (1977).

This is not to say that the techniques are always correctly applied. Some common violations are discussed below.

8.2. Violations of Independence

Diaconis (1978) has criticized the use of trial-by-trial feedback with a "closed" ESP target sequence. With feedback, subjects can make inferences about the remaining targets in the sequence, and obtain spuriously high scores. This has been an issue in remote viewing research (see Section 4.7). However, Kelly (1979) has argued that this is not a frequent source of error in ESP experiments, since most parapsychologists are aware of the danger. In support of Kelly I found no such cases within my sample (though the target deck used by Johnson & Kanthamani, 1967, was partially closed). I found that where closed-deck sequences had been used, there was no direct trial-by-trial feedback. However, there were two cases in the sample where subjects could have obtained such feedback through conversations with fellow subjects (Braud & Braud, 1973, Exp. 2; McBain et al., 1970). This may also have been a problem in the Moss et

al. (1970) study, if control target slides were sometimes borrowed from "episodes" that the subject's testing partner had already seen. (The description of the experiment is unclear on this point.) There was a similar possibility for indirect feedback in experiments by Honorton and Harper (1974) and Terry and Honorton (1976). However, in the latter studies any feedback effect would be too small to account for the observed results.

Diaconis (1978) also refers to the problems that can arise when multiple responses to a single target are treated as if they were independent. Suppose, for example, that "flower" is always the first target in a free-response experiment, and that this corresponds with the response preference of the subjects. There may be an enormous number of "hits" on the first trial, but this is obviously no evidence of ESP. This is only evidence that the subjects agree among themselves. Goodfellow (1938) made this argument long ago, in a critique of the Zenith Radio mass-broadcast ESP tests. Often, a similar problem has arisen in remote-viewing research, where multiple judgments of a transcript/target pair have been treated as if they were independent. This error has been discussed by Child and Levi (1980), Marks (1982), and Stokes (1980).

Multiple-call experiments can be appropriately analyzed by applying either a majority-vote method, or the more sensitive "Greville technique" (Burdick & Kelly, 1977; Greville, 1944). Within my 54-experiment sample, there were five multiple-call experiments (Bhadra, 1966; McBain et al., 1970; Musso, 1965; Schmeidler & McConnell, 1958, Group Exps.; Wilson, 1964, Exp. 1). Surprisingly, in none of these cases were the data analyzed appropriately. Musso (1965) realized the problem and "corrected" his z-score—but in a manner that Burdick and Kelly (1977, Footnote 5) have rejected as inappropriate.

The multiple-calling problem in Schmeidler's research was pointed out by Scott (1951a). In response to Scott, Schmeidler (1951) acknowledged the difficulty, but questioned whether it was a matter of practical concern in such large-scale experiments. Presumably, the dependence among subjects' calls is rather small in an experiment with so many target lists. Scott (1951b) saw some logic to Schmeidler's argument and withdrew his criticism (though he later found other reasons to question Schmeidler's results, as discussed in Section 7.2). Empirical comparisons (Davis, 1978; Humphrey, 1949) confirm that in multiple-call data from typical forced-choice ESP tests, it makes little difference whether the results are analyzed by the usual binomial formula, or by the more appropriate Greville method. For this reason, I decided to assign a flaw only in the case of a study by McBain et al., 1970, where the results were marginally significant. Hopefully, these data are still available and can be reanalyzed.

It should be noted that the situation is quite different in remote-viewing experiments, where the number of targets is generally small. It then becomes critical to control for response bias (Marks, 1982). In the remote-viewing experiments, violations of independence have arisen even where there was only one response per target. In several such studies a single judge has evaluated the correspondences between targets and responses. The judge may, under these conditions, be influenced in assigning a rank or rating to a given target by the mem-

ory of how he or she assigned ranks or ratings to other targets (Kennedy, 1979a, b; Morris, 1972; Scott, 1972). The data from such experiments can still be analyzed, but it is necessary to apply the method described by Scott (1972). Scott's (1972) article does not seem to have been widely read, since researchers have often made false assumptions of independence, and applied a binomial analysis. Kennedy (1979a,b) has cited such cases, three of which were in my sample (Honorton, 1972; Honorton & Stump, 1969; Parker & Beloff, 1970, Exp. 1). Kennedy's reanalyses show that Honorton's (1972) results are actually nonsignificant, while results from the other two studies would remain significant when maximum dependence was assumed.

Questions of statistical dependence can also be raised with respect to a hypnosis study by Krippner (1968). In this study judges' ratings were entered into matrices of targets by transcripts, and evaluated by an analysis of variance. However, entries in such a matrix cannot be assumed to be independent. As Kennedy (1979a, Footnote 4) observes, it is unclear how the analyses were conducted, or how the dependencies were taken into account. Until these data (which may no longer be available) can be reanalyzed, their significance will remain in doubt.

In summary of results from the sample, there were five studies with possible violations of independence (Braud & Braud, 1973, Exp. 2; Honorton, 1972; Krippner, 1968; McBain et al., 1970; Moss et al., 1970). The studies by Honorton, Krippner, and McBain et al. were not previously cited for flaws. Hence their inclusion increases the proportion of flawed experiments from 40/54 to 43/54, in the overall sample.

8.3. Optional Stopping

In Rhine's (1934/1964) early ESP experiments there was generally no prespecification of the number of trials to be conducted. Did the experimenters stop a series of trials at an optimum time, just when "significance" was achieved? Lemmon (1937) and Leuba (1938) suggested that this had in fact happened, and that Rhine's results were, at least in part, artifacts of optional stopping. To answer this question, Greenwood (1938) examined a lengthy control series (500,000 mock ESP trials), which he cut into subseries of varying lengths. He found that within the range of trials typically encountered in ESP studies, the optional stopping effect was not as much of a concern as these critics had imagined, and that it could not explain Rhine's extraordinary results. Greenwood concluded that optional stopping was a legitimate criticism only in those studies where the results were marginally significant.

Among the studies in my sample, there were many with marginally significant results. Hence, the possibility of optional stopping had to be considered. I found that the number of trials per subject had nearly always been preset. However, there was often no indication of whether the number of *subjects* had been preset. This number was usually a multiple of 10, which at least suggests a pre-

set figure. I decided to assume that the number had been preset, unless there was some reason for doubt. I was especially concerned in cases where an experimenter obtained intermittent feedback and stopped the experiment at a time when marginally significant results had been achieved. There were two such cases in the sample (Honorton & Stump, 1969; Stanford & Mayer, 1974). In the former case, the experiment was stopped when the two experimenters were suddenly forced to relocate. In the latter case, the experiment was stopped so that the experimenters could prepare a convention report. Since the stopping point was less well-defined in the latter case (Stanford & Mayer, 1974) I have chosen to assess a flaw; it is difficult to exclude the possibility that Stanford and Mayer were unconsciously influenced by the trend of the results (which yielded a p-value of .03, one-tailed). The Stanford and Mayer study, which was generally well-controlled, also allowed a minor possibility for handling cues (see Section 4.9).

8.4. Data Selection

If "poor" data are thrown away, the spurious effects can be much larger than those arising from optional stopping. Critics of Rhine often accused him of discarding low-scoring runs (Pratt et al., 1940/1966, pp. 414-415). In his early research thousands of trials fell into the category of "preliminary" or "practice" runs which Rhine chose to exclude (Rhine, 1934/1964, pp. 76-77). Initially, the distinction between these runs and "record" runs may have been informal, but it appears to have been formalized in the last year and a half of the research (Rhine, 1934/1964, p. 76). Data selection may still have played a role, but it cannot entirely account for results such as those achieved by Hubert Pearce, which were nearly twice chance expectation.

Data selection has continued to be an issue in parapsychological research. Gardner (1975, 1981) considered this a plausible explanation for "Phase 1" results from an ESP learning study, conducted by Targ, Cole, and Puthoff (1974). Gardner claimed that the subjects (who worked unobserved in Phase 1), had turned in their paper tape records in "disconnected bits and pieces" (Gardner, 1975, p. 116). Targ and Puthoff (1976) denied this, maintaining that "the tapes were always delivered to us intact with all runs recorded" (p. 6). However, irrespective of whether the tapes were in one piece, there was apparently no assurance that portions had not been torn from the beginnings or ends of the tapes (Hansel, 1980, pp. 240-241). This hypothesis does not necessarily imply cheating. The subjects may have felt it was legitimate to discard data, if they misunderstood the requirements of scientific method.

In most parapsychological research, subjects have no opportunity to select data. However, the experimenters usually have this opportunity. Marks and Kammann (1980) suggested that Targ and Puthoff (1977) were guilty of data selection in their "remote viewing" research: "If visitors came to try remote viewing for themselves and the results were good, these were 'experiments.' If

the results were not so good, they were 'demonstrations' " (p. 38). Marks and Kammann provided no firm evidence of selection, and were forced to withdraw one such allegation (Marks, 1981a). They did, however, provide evidence of selection in Targ and Puthoff's (1974) ESP studies with Uri Geller. Comparing the 1974 report with a "daily log" of the same experiments (Puthoff & Targ, 1976), Marks and Kammann found discrepancies with respect to (a) whether a trial had been designated a "pass" or a "real attempt," and (b) whether completed drawings were or were not submitted to judges. The ambiguities suggest possibilities for an accidental selection bias. It may be that Targ and Puthoff can answer this allegation, by distinguishing between two varieties of "passes," only one of which was to be excluded from the results. However, they have not yet published a formal rebuttal.

8.5. Data Selection within the Sample

In my 54-experiment sample, there were four cases where data selection may have contributed to the observed scoring rates:

1. Terry and Honorton (1976, Exp. 2): This was a 10-session experiment. If subjects quit the experiment before completing the 10 sessions, their data were not retained. However, as Kennedy (1979a) noted, these may well have been the subjects who were scoring low. Their data ought to have been included.

2. The same question can be raised with respect to a study by Braud and Wood (1977), which required subjects to make six visits to the laboratory. The authors make no mention of subjects who failed to complete the six sessions. However, it is more than likely that there were a few such persons, since the subject pool consisted of unpaid visitors to the lab. Until this point can be clarified, the study will appear to be flawed.

3. Parker and Beloff (1970, Exp. 1) conducted a hypnosis study with two test sessions per subject. Overall, results were nonsignificant, with a deviation of only one hit above chance. The authors did an analysis of first session results, which were significant. Although the success of the first session is worth noting, the study as a whole was clearly a failure to confirm, and should not have been cited by Honorton (1977) among his "successful" hypnosis studies.

4. Moss et al. (1970) included in their "hypnosis" group a percipient who had at the last minute refused to be hypnotized. The grounds for including these results were that the agent was hypnotized. However, this was "contrary to original design" (p. 53) and not in line with previous research. Moreover, the significant scoring by this percipient was described as "unexpected" (p. 54). There were, therefore, clear grounds for excluding the data of this percipient (who obtained the highest overall scores). If the data of this subject were excluded, the results for the hypnosis group would apparently remain at least marginally significant. Hence I did not assign a flaw in this borderline case.

Of the first three studies, the ones by Parker and Beloff and by Braud and Wood had not been cited in any previous sections. Their inclusion increases the overall proportion of flawed experiments from 43/54 to 45/54.

8.6. Multiple Analysis

After the publication of Rhine's (1934/1964) *Extra-sensory Perception*, there was a substantial decline in the ESP scores achieved by Rhine and his staff. Pope and Pratt (1942) attributed this decline to the increasing use of unselected subjects, and to the formality of new testing techniques, which took away from "spontaneous interest" (p. 14). Critics were naturally inclined to suppose that it was the increased experimental safeguards that accounted for the decline of scores. At any rate, as the decline set in Rhine and his colleagues turned increasingly towards analysis of secondary effects, such as position effects, displacement scores, etc. Overall scoring was often at chance, yet evidence for ESP was being claimed on new grounds. This strategy did not sit well with Rhine's critics, one of whom questioned "the advisability of studying the detailed ramifications of an alleged phenomenon before the phenomenon itself has been incontrovertibly demonstrated" (Foley, in Sells, 1941, p. 252). Rhine held to the position that the secondary effects, such as declines, could by themselves constitute evidence for ESP.

With so many ways of analyzing ESP data, there arose the question of whether significance levels were inflated by multiple analysis. This issue has continued to be raised, not just by critics, but by parapsychologists themselves. Wicklund (1977), for example, suggested that many reported declines, sex differences, and other secondary psi effects, were simply artifacts of *post hoc* analysis. Kennedy (1979a,b) was concerned about the wide variety of ways in which free-response experiments had been analyzed; he concluded that the *p*-values claimed for these experiments were often inflated. Honorton (1979) saw a greater problem in fixed-choice ESP experiments, where results have been analyzed in terms of "*post-hoc* decline effects, *n*th level forward and/or backward displacement, psi-missing, variance effects, time of day, relationship to x-variable, etc." (p. 383).

On the critic's side, Hyman (1983) has surveyed the effects of multiple analyses on claimed replication rates for ganzfeld ESP experiments. He concludes that, just by chance, one might expect as many as 25% of such experiments to yield "significance" (at the apparent .05-level), if the effects of multiple analysis were not taken into account. Hyman, along with Kennedy, argues that it is difficult to know how many analyses were conducted, and how many more would have been conducted if significant results had not been obtained. In his survey of 42 published ganzfeld experiments, Hyman claimed to have found only three cases where investigators had reported a single, preplanned method of analysis.

I will not comment in any detail on Hyman's survey, since his complete analysis has yet to be published, at the time I am writing. However, I believe the problem may not be quite so serious as it initially appears. There are many instances where one can infer prior planning of a statistical analysis, even though the analysis is not described as preplanned. Often, for example, an investigator applies a technique which he or she has consistently applied in previous research. It then makes sense to assume that the analysis was planned. Honorton (1979, 1983) makes two further arguments in response to Kennedy and Hyman: He

observes, first, that many of the ganzfeld experiments have been significant on a straightforward binary or direct hit analysis. Second, the obtained p-values in many such studies are low enough to survive corrections for multiple analysis.

In my sample, there were 11 ganzfeld experiments which had been published in full. (Hyman's survey included many convention papers or other minor publications.) Among the 11 studies, the results were significant on a direct-hit analysis in seven cases, and on a binary-hit analysis in two cases. There was one study (Sargent, Bartlett, & Moss, 1982) where significance was claimed on the basis of a preplanned sum-of-ranks test. In the remaining study (Braud & Wood, 1977) there were, as Hyman notes, a bewildering variety of analyses. The investigators used the Maimonides slides as their target material, and on that basis, one can assume that they intended to apply the binary coding system which Honorton devised for the slides. I am impressed by the results of Table 3 (p. 420), on the basis of which I compute an overall t of 3.88, with $df = 29$, for an analysis based on binary coding. Since this result would withstand corrections for multiple analysis, I do not believe that a flaw should be assigned.

Both Hyman (1983) and Sargent (1980b) have questioned whether the binary hit analysis in Braud, Wood, and Braud's (1975) study was preplanned. Stanford (1981d) argues that it was, since the Brauds used the same analysis in previous research (e.g., Braud & Braud, 1973; Braud & Braud, 1974). In the ganzfeld condition from Braud, Wood, and Braud, the investigators obtained 10 binary hits and 0 misses, a result which would again withstand some correction for multiple analysis. The only other analysis which the Brauds had used was an analysis for direct hits, as in their earlier research. On that basis, I saw no reason to assign a flaw.

Finally, mention should be made of the binary hit analysis used by Terry and Honorton (1976, Exp. 1). Kennedy (1979a) raised some legitimate questions as to whether the analysis had been preplanned. In his previous ganzfeld research, Honorton had used only a direct-hit analysis (Honorton & Harper, 1974). If direct hits had been the measure of significance in the Terry and Honorton (1976, Exp. 1) study, it would have been deemed nonsignificant. (The direct hit rate just misses significance, with $p = .053$, one-tailed.) In a further study Terry and Honorton (1976, Exp. 2) reverted to a direct-hit analysis. Yet, Honorton (1979) has provided strong assurances that the binary-hit measure was preplanned, and on that basis I chose not to assign a flaw. However, it would have been preferable if these assurances had been included in the original experimental report.

Later, investigators began using measures which ought to be more sensitive than simply counting the number of hits. Sargent (1980b) has used a sum-of-ranks analysis (in addition to counts of direct hits). Palmer (e.g., Palmer & Vassar, 1974) and Stanford (e.g., Stanford & Mayer, 1974) have used measures based on subject ratings. Obviously, the danger of overanalysis does exist. Yet, so long as each investigator prespecifies a method, the error rate remains at alpha, even though this may differ from what he or she used previously, or from what others are using. It is not surprising that investigators differ in their meth-

ods, since all of the free-response research is in its preliminary stages; there are disagreements on many aspects of the experimental designs.

8.7. Multiple Analysis within the Sample

Among the experiments in the sample, there were many where the authors had failed to state that an analysis was preplanned. Usually, however, I could infer that the analysis had been planned (based, for example, on the authors' previous research). I did assign a flaw of multiple analysis in studies by Braud and Braud (1974, Exp. 1) and Grela (1945), as explained below.

In the experiment by Braud and Braud (1974, Exp. 1) the results were marginally significant ($p < .05$, one-tailed) on a binary hit analysis, while they were nonsignificant on the basis of direct hits. In their previous research (Braud & Braud, 1973), the authors had given equal emphasis to the binary and direct hit analyses. If direct hits were a potential analysis in the later study, as well, then a correction is in order. With a correction for multiple analysis, the results are nonsignificant.

In the experiment by Grela (1945) subjects were tested under conditions of positive and negative suggestion, in an attempt to reproduce Schmeidler's finding of a scoring difference dependent on attitudes towards ESP (see Section 7.2). In the condition of positive suggestion, subjects were told that "telepathy is real" (p. 198). In the condition of negative suggestion, they were told that they "did not believe in telepathy" (p. 199). Hence the logical analysis was an assessment of the *difference* between means in the positive and negative conditions. This would be analogous to Schmeidler's test of a difference between sheep and goats (Schmeidler & McConnell, 1958). Yet, that difference is nonsignificant in Grela's research. Grela claimed significance by looking at the *sum* of deviations across all conditions, which yielded a z-score of 3.43. This is an impressive z, but it may have been inflated by an inadvertent selection of data: Presumably, some subjects were initially in the "control" session, but did not continue on to complete the remainder of the experiment. Many of these subjects may have left the experiment because they obtained low ESP scores, and were discouraged.

Since these two experiments were cited for flaws in previous sections, the overall proportion of flawed experiments remains at 45/54.

9. Reporting Failures

Many of the sample studies have already been cited for failures to describe such procedural details as randomization or recording. Nonetheless, the sample studies as a whole were well-described, at least in comparison with Rhine's (1934/1964) early ESP studies. Description of Rhine's initial work was so bare that no assessment was possible. Dingwall (1937) concluded that "the experimenters have but slight idea of the kind of report which is necessary for scien-

tific men to be able to understand the work being accomplished" (pp. 140-141). In response to a similar assessment by Hansel (1966, 1980), Rao (1981) has defended Rhine, arguing that "details of the sort that we now require . . . were simply not found necessary then" (p. 192). There is some truth in Rao's assessment. It should be emphasized that Rhine was not attempting to "prove ESP" in the early research. He thought that there was already ample evidence for the phenomenon (Rhine, 1934/1964, Chapter 2). It was only in the late 1930s, as the research did become more proof-oriented, that reporting standards improved (e.g., Pratt & Woodruff, 1939).

9.1. Targ and Puthoff's Nature *Article*

Modern-day parapsychologists tend to underestimate, as did Rhine, the amount of detail necessary to convince skeptical scientists. A case in point is Targ and Puthoff's (1974) report on Uri Geller. *Nature* had published the article despite the opinion of three referees that details of the experimental procedure had not been adequately described (see the editorial accompanying the article). Marks and Kammann (1980) took note of the reporting omissions, and speculated on whether damaging facts might have been left out. They concluded that the actual conditions may have allowed cheating by Geller.

Morris (1980b) agreed that reports of the Geller experiments were inadequate. However, he attributed this in part to space limitations imposed by *Nature.* In the same article Morris noted that an author always has the option of providing additional details in unpublished documentation, which is then made available. Morris advocates keeping an open mind on the Geller experiments, since new information from the authors may yet rule out the scenarios for cheating. I would agree, but it is much preferable if such information is made available at the time of the experiment, rather than years later. (See also Marks, 1981; Morris, 1981; Targ & Puthoff, 1977, pp. 166-188.)

Gardner (1982-1983) tried to obtain more information on one of the reported experiments, in which Geller attempted to guess the uppermost face of a hidden die. Geller made eight guesses, and was correct on all eight attempts (Targ & Puthoff, 1974). The result was unquestionably significant, but details were lacking, such as the exact dates of the trials, names of witnesses present, and especially information on whether Geller was permitted to touch the die container. Gardner claims that he was unable to obtain these details, and that the "experiment" had no more than anecdotal status. Until Gardner's questions are answered, this conclusion will hold. Hyman (1976) has reviewed various research with Geller, and found that this anecdotal style of reporting is quite typical of work with that subject.

Since Geller has not made himself available for subsequent laboratory study, the reported experiments cannot be repeated. Morris (1980b) correctly sees repeatability as the key to the controversy. If an experiment cannot be repeated, it is always possible for a critic to suppose that some procedure was not correct-

ly followed, or that there was a critical reporting omission. For that reason, Morris believes that isolated experiments can never provide evidence for ESP; he argues that the evidence can only come from repeatable experiments, such as those discussed in the Wolman *Handbook* and in volumes of the present series.

Earlier, Crumbaugh (1966) made a similar argument. He noted that what an experimenter has actually done "may deviate in unrecognized ways from what he thinks he has done" (p. 52). Hence, only a repeatable experiment can be adjudged as proof of ESP. However, Crumbaugh called for total repeatability. Morris argues that repeatability need only be clearly above the chance level, and sufficient for experimental progress. I agree with Morris that critics should not insist on total repeatability at this early stage of parapsychological research. If the repeatability is low, however, then critics can reasonably insist on unusually high standards for the experimental reports. This is especially the case if "low repeatability" means that some experimenters can never elicit the effects, even under "psi conducive" conditions. (Hopefully, the repeatability problem is not quite this severe.)

9.2. Reporting Failures in the Sample

Deficiencies in reports from my sample have already been discussed, for the most part, under appropriate headings. In some cases these deficiencies or omissions appear to indicate a lack of planning. If an experiment has not been planned, and there was no written protocol, then our knowledge of how it was conducted depends entirely on the investigator's ability to accurately recall what he or she did. This recalled procedure is likely to differ from the actual procedure, since it is based only on human testimony (e.g., Loftus, 1979).

On the other hand, a reporting omission does not always imply lack of planning. A ganzfeld researcher might, for example, focus so closely on details of the ganzfeld induction, seeing this as crucial to a successful replication, that he or she forgets to report on the method of target randomization. The randomization procedure might have been carefully planned, but simply left out of the experimental report. Keeping this possibility in mind, I do encourage a further "information exchange" with authors, as Morris (1980b) advocates. It is preferable if such exchanges take place in print, rather than in private correspondence; otherwise, there may be no public benefit. In the controversy surrounding Targ and Puthoff's (1974) research, there has been a voluminous private correspondence (see Marks, 1981a; Morris, 1980b, 1981) which apparently has left several key questions unresolved.

It is the responsibility of scientists to provide most of the relevant details in their initial reports. A seriously flawed experimental report cannot be salvaged by "information exchange." Among the studies in my sample, procedural details were especially lacking in studies by Haraldsson (1978); Honorton and Harper (1974); Musso (1965); Rao, Dukhan, and Rao (1978); Shields (1962); and

Schmeidler and McConnell (1958). Rao, Dukhan, and Rao (1978) wrote a report of a large study (92,600 card trials) that contains virtually no details on recording and safeguarding of data. Musso (1965) conducted another very large experiment, in which 302 subjects were tested by 11 different experimenters. Yet, the description of the procedure occupies less than one page of text. Haraldsson conducted a smaller study, but he used 14 undergraduate experimenters; details were lacking on their training and/or supervision, nor was there a description of their ESP test machine. Honorton and Harper (1974) failed to clarify the locations of their agents, percipients, and experimenters. Shields (1962) provided almost no procedural details in a report on personality correlates of ESP. Schmeidler and McConnell (1958) described "typical" sessions, but did not explain changes in the sheep-goat criteria from one session to the next. These details are rarely available from earlier reports of the same experiments (see Section 7.2).

With the exception of the study by Musso (1965), these experiments have already been critiqued under previous subheadings. The inclusion of Musso (1965) increases the proportion of flawed experiments, in the sample as a whole, from 45/54 to 46/54.

10. Experimenter Fraud

There were no experiments within the sample where I had reason to suspect experimenter fraud. On the other hand, none of the experiments included strong controls against fraud. It would not be surprising if there were a few instances of fudged or "tidied" data, though outright faking of data is presumably rare.

Data faking was a special concern of J.B. Rhine, who often wrote on the issue. Rhine (1974b) once described 12 cases of fraud within parapsychology, from the 1940s and 1950s. None of the individuals was identified, although four had been caught "red-handed" (p. 104). As Nicol (1976) observes, there is no way of knowing the number of publications in which these four individuals were involved.

Within the same article, Rhine discussed the current scene, concluding that "we have been able to do quite a lot to insure that it is impossible for dishonesty to be implemented inside the well-organized psi laboratory today" (p. 105). Ironically, Rhine (1974a) discovered a new case of fraud only a few months after his article appeared in print. Not only was the case within Rhine's own laboratory, it involved Levy, the acting director. Evidence emerged that Levy had been faking data for over a year prior to being caught (Rhine, 1975).

The fraud issue was raised at a much earlier date, by Price (1955). When Price raised the issue he was accused by both parapsychologists and orthodox scientists of having made wild, unsubstantiated charges (see the correspondence in subsequent issues of *Science*). This perhaps illustrates Rosenthal's (1976)

dictum that "the charge of fraud . . . is leveled only at the peril of the accuser" (p. 33). Price had focused on the research of Rhine in America and Soal in England. He later corresponded with these investigators, and was influenced to withdraw the allegations (Price, 1972). Yet, evidence accumulated in later years that Soal may in fact have been guilty of data fabrication (Hansel, 1980; Markwick, 1978; Scott & Haskell, 1974). Markwick's argument seems fairly conclusive, but her supporting analyses are involved, and have not yet been independently checked.

Experimenter fraud has frequently been raised as a counter-hypothesis to psi (e.g., Girden, 1978; Hansel, 1966, 1980; Markwick, 1978; Medhurst & Scott, 1974; Price, 1955; Scott & Haskell, 1974). Yet the actual incidence of fraud is unknown. Broad and Wade (1982) reviewed many actual or apparent instances of scientific fraud, and concluded, "fraud is endemic" (p. 224). Their review does at least establish that fraud is much more common in science than had previously been supposed. If faking of data is common, how much more common are "lesser" violations of scientific ethics, e.g., data "massage," data suppression, distortions in experimental reports, and the use of "hired-hand" assistants (Roth, 1966) who approach their experimental tasks with less than the most lofty motives? These "lesser" violations generally pose little risk to senior investigators. If they are found out, they can generally plead guilty to carelessness; there will be no evidence of deliberate deception.

The general opinion seems to be that fraud or fudging cannot be completely controlled. Since it cannot be controlled for, it will always represent a "last-ditch" alternative for the skeptic, when he or she is presented with strong evidence for psi. In short, experimenter fraud is an unfalsifiable hypothesis (e.g., Honorton, 1981; Morris, 1980b).

I would agree that experimenter fraud is difficult to control for, but I do not agree that controls are impossible. Johnson (1975) and Schmidt (1980) have proposed models which I believe would satisfy all but the most hardened skeptic, on this point. The models are too complex to be more than briefly mentioned here. In Johnson's "Model 3" the experimenter cannot fake data, because he or she does not have access to the targets until after the subject's calls have been recorded by a distant computer center. In Schmidt's model, fraud is eliminated because the experimenter does not know, at the time the data are collected, which trials are experimental and which are controls.

The ease with which such methods can be introduced depends on the nature of the experimental design. It might be difficult to introduce these controls into a telepathy experiment, for example. However, the choice is not between complete control and no control at all. There is some value in introducing partial controls against fraud. The situation is analogous to classroom cheating. Certainly, the standard procedure of monitoring students during an exam does not prevent cheating. A determined trickster might, for example, use a miniature electronic receiver in place of crib sheets. Yet monitoring of the classroom does cut down on fraud, simply because it makes fraud more difficult. Presumably, scientists as well as students cheat, in many cases, simply because cheating is

easy (as in a take-home exam). To cut down on cheating, it is not necessary to have a fraud-proof environment; it is only necessary to have a realistic threat of fraud detection.

Among the experiments in my sample, there were several which introduced such partial controls (Haraldsson, 1978; Kanthamani & Rao, 1971, Exp. C; Parker & Beloff, 1970). In the study by Haraldsson, a precognition task was administered to subjects in Iceland. Targets were then generated by Schouten, at the University of Utrecht in Holland. Schouten mailed the targets to Haraldsson on the same day that Haraldsson mailed the calls to Schouten. Under these conditions, fraud may still be possible, but it is at least more difficult. (Unfortunately, there was no significant scoring under these conditions.)

In Experiment C by Kanthamani and Rao (1971), there was a similar exchange of data. ESP and the personality data were collected by different experimenters. At a later time the experimenters exchanged their data and computed the correlations between ESP and personality. Though details on the exchange are lacking, it presumably would reduce the possibility of inadvertent bias, and perhaps deliberate fraud as well. Some significant correlations were obtained.

Finally, Parker and Beloff (1970) included modest controls against fraud in a free-response hypnosis experiment. For each session, there were four targets (art reproductions), which were sealed into their envelopes by a person from outside the laboratory. The set of envelopes was then shuffled further and labeled by the main experimenter (Parker). During the experiment, duplicate sets of subjects' responses were given to the outsider, who was present at the conclusion of the experiment for an unsealing of the envelopes. If the envelopes had been carefully sealed (details are lacking), then it would have been difficult for Parker to have faked results on his own. The question is academic, however, since the overall results were "not sufficient to demonstrate clairvoyance" (p. 436). (See also Section 8.5.)

Some parapsychologists have concluded that controls such as these are unnecessary. Palmer (1978) sees the possibility of a "climate of paranoia" in which researchers could not function. McConnell (1977) concludes that "a certain amount of risk must be tolerated if science is to progress in its traditional fashion" (p. 433). The proposed alternative is an increased emphasis on independent replication (e.g., Honorton, 1981; Palmer, 1978). If psi results are obtained by enough independent investigators, then the fraud issue will disappear.

However, this sole reliance on replication rates is somewhat of a gamble. If replication rates are high, then the fraud issue is indeed settled (assuming that the experiments are sound in all methodological aspects). Suppose, on the other hand, that replication rates are low, as they have tended to be in the past. Then experimenter fraud will remain a viable explanation. Yet, the low replication rates might arise from uncontrolled variations in a genuine psi effect, and it would be decades before this became apparent.

11. Summary

11.1. Conclusions from the Sample

Results from the 54-experiment survey have demonstrated that there are many alternative explanations for ESP phenomena; the choice is not simply between psi and experimenter fraud. There are numerous alternatives, because experiments are often inadequately controlled, or they are reported too loosely for any such evaluation to be made. The numbers of experiments I judged to be flawed on various grounds were as follows: randomization failures (13), sensory leakage (22), subject cheating (12), recording errors (10), classification or scoring errors (9), statistical errors (11), reporting failures (10). (The sum of these numbers surpasses 54, since experiments were often flawed on more than one basis.) All told, 85% of the experiments were considered flawed (46/54).

This leaves eight experiments where no flaws were assigned (Bhadra, 1966; Braud & Mellen, 1979; Honorton & Stump, 1969; Kanthamani & Rao, 1971, Exp. C; Sargent, 1980b, Exps. 2, 3, & 5; Sargent, Bartlett, & Moss, 1982). Although none of these experiments has a glaring weakness, this does not mean that they are especially strong in either their methods or their results. Weaknesses in the eight experiments are discussed below:

Bhadra (1966): This was a borderline study, in which I almost assigned a flaw for classification error (see Section 7.3). The subjects were classified as "sheep" and "goats" only after the data were collected. Because the basis for classification seemed a reasonable one, I did not fault this procedure. However, the design clearly does not satisfy the criteria set by Hansel (1980, p. 202) for a study of this type because the classification was not preplanned.

Braud and Mellen (1979): This was a hypnosis experiment in which subjects were "age regressed" to see whether this might enhance their ESP scores. The "age regression" did not enhance scoring, but there were, at least, significant overall results in the hypnosis condition, with $t\ (9) = 2.73, p = .015$, one-tailed. When the hypnosis results were compared with a nonhypnosis condition, the difference was nonsignificant. Hence, there is actually no support for the use of hypnosis as a "psi-conducive" procedure. A minor weakness in the procedure was the placing of the target in the subject's lap (though the target envelope was stapled shut). There was also no description of the scoring procedure, though there would seem to have been little possibility for error.

Honorton and Stump (1969): The results of this hypnosis study, as reassessed by Kennedy (1979a, p. 4), are significant at approximately the .05-level, two-tailed. If a correction was made for multiple analysis (judge's evaluations were nonsignificant), the results would probably remain significant at the .05-level, one-tailed. Randomization of the targets was by informal means (shuffling) and the experiment was stopped before the preplanned number of sessions had been completed. I did not assign a flaw (Section 8.3), because there seemed to be legitimate reasons for the investigators' having stopped. A minor weakness in

the procedure was the placing of an (unsealed?) target envelope in close proximity to the subject (who was, however, not permitted to handle it).

Kanthamani and Rao (1971, Exp. C): This was a study of personality correlates of ESP. The description of the procedure is skeletal, though further details are available in Kanthamani's (1969) dissertation. Again, randomization was informal (shuffling). One-half of the ESP data were collected by Rao, an experienced investigator. However, the other half were collected by an experimental assistant who was relatively inexperienced. It is unclear whether the personality differences would remain significant if Rao's data were considered separately.

Sargent (1980b, Exps. 2, 3, & 5); Sargent, Bartlett & Moss (1982): These four ganzfeld studies are considered together, since they all employed a similar methodology. The weakest of these was Sargent's (1980b) second experiment, where the results were only marginally significant, and where there was some question as to what defined the beginning and end of the experiment. This experiment overlapped in time with Sargent's (1980b) first experiment. In fact, the first four sessions of the second experiment could, it would seem, have equally well been assigned to the first experiment.

The Sargent experiments seem to have been well-controlled, on the whole. Stanford (1981d) notes that there is a certain vagueness in Sargent's descriptions of his randomization procedures; this was the case in two of the sample studies (Sargent, 1980b, Exps. 3 & 5). I would assume that the procedure was similar to that in his initial experiments.

My main concern, in the Sargent experiments, is with possibilities for subject cheating. In some trials (Sargent does not say how many) the agent is someone other than an experimenter. This agent is in a shielded room, during part of the experiment, but the room is not electrically shielded. Moreover, the agent leaves the shielded room before there is any assurance that the percipient has completed all the rankings. Perhaps the agent would have an opportunity, after leaving the shielded room, to pass information to a confederate, who would then signal the percipient. An argument against that hypothesis is that Sargent has obtained significant results with naive subjects (Stanford, 1981d). On the other hand, his strongest results have been with experienced subjects who have had a close relationship with the agent.

In conclusion, eight experiments were conducted with reasonable care, but none of these could be considered as methodologically ideal. When all 54 experiments are considered, it can be stated that the research methods are too weak to establish the existence of a paranormal phenomenon.

11.2. On Salvaging Flawed Experiments

Honorton (1981) argued that all experiments have "potential flaws." Rather than reject experiments on this basis, Honorton suggests that critics attempt to empirically evaluate the status of the flaw. He outlines two alternative courses of action: (a) critics can compare flawed and unflawed samples, to see whether

the psi scoring rates differ; or (b) they can "repeat the experiments, systematically varying the presence or absence of the suspected flaw" (p. 159). If psi scoring rates are the same in flawed and unflawed samples, or if the experimental introduction of the flaw makes no difference, then the "flaw" can be seen to have no real-life consequences.

Honorton (1979) applied the first approach to the issue of handling cues (which were discussed in Sections 4.8 and 4.9). He found that ESP scoring rates were similar in studies which allowed, or did not allow handling cues. Thus, it apparently made no difference whether or not this variable was controlled for. Honorton's implicit assumption is that the two samples were comparable in all respects, except for the presence or absence of the handling cue possibility. However, this assumption is one that most skeptical outsiders would be unwilling to make. If a study is "unflawed" with respect to handling cues, and yielded strong evidence of ESP, then the natural suspicion of the skeptic is that the study may be flawed in some other respect, such as randomization or recording. The skeptic may be wrong, but certainly this possibility must be taken into account. Hence, as Kennedy (1979b) and Hansel (1980) have argued, the only meaningful analysis is one in which all the methodological variables are considered together.

Other variables can be held constant only if the presence or absence of the suspected flaw is experimentally manipulated (Honorton's second alternative). This is a more viable choice, but it presents practical difficulties. To do this, one needs to know the precise nature of the flaw (e.g., what the handling cues were). This information is generally unavailable, because the flaw resulted from uncontrolled variables. These variables were never measured. For this reason, they are usually too poorly defined to allow the experimental manipulation that Honorton desires. In this sense, a "potential flaw" is worse than an "actual flaw" (whose effects have been identified and can be objectively assessed).

An analogy can again be made to the proverbial "dirty test tube" (Section 3.5). No one knows whether the dirt introduced an artifact, because the nature of the dirt is unknown (or it would not have been allowed in the test tube to begin with). The dirty test tube represents no more than a "potential artifact." Nevertheless, the investigators have no choice but to repeat the experiment with a clean test tube. The results cannot be salvaged.

Since critics of parapsychological research identify "potential flaws," they do not usually succeed in establishing an alternative explanation for psi effects. According to Honorton (1975a, 1979) their case is thereby weakened. However, the critics are not ordinarily pushing for acceptance of an alternative hypothesis. They are usually asking only that claims for psi be suspended until properly controlled studies can be carried out. The burden of proof lies not with the critics, but with the parapsychologists. That situation is not peculiar to parapsychology. In science generally, new claims (and especially startling new claims such as a cure for cancer) must withstand the critical scrutiny of skeptical scientists.

On the other hand, there is a danger that critics will overstep their bounds, and engage in speculative criticism long after the major methodological problems

have been solved. Honorton (1975a, 1979) believes that this is already the situation in parapsychology. However, the present survey suggests that there are methodological issues still to be solved, and that the critics have something positive to offer. Indeed, I feel that critics have a responsibility to do more than just criticize; they must offer constructive suggestions, and work with parapsychologists in the design of new experiments, which exclude all counterhypotheses.

11.3. Methodological Recommendations

To provide a list of recommendations would be only to reiterate points already made in previous sections. Instead, I will offer a more general suggestion—that parapsychologists attend more seriously to possibilities for psi artifact. I find that sources of artifact are rarely discussed in the sample studies, where ESP is considered as a well-established process. This attitude has been explained by Rao (1979) as follows:

> It might be appropriate . . . to clarify a situation that confuses many people. Those individuals who are neutral or skeptical about the existence of psi are primarily concerned about Type I errors and thus are interested only in the rigorous evidential value of each work. On the other hand, those of us who find the evidence convincing are more concerned with learning about the relevant variables. This leads us to be more concerned than skeptics about Type II errors and thus to place more weight on work that may not be of the highest standards from an evidential point of view [p. 417].

Parapsychologists such as Rao, who accept the existence of psi, argue that parapsychology should be evaluated on much the same basis as other behavioral sciences. They see the evidence for psi as already sufficient; hence they see no need for the unusual precautions taken in "proof-oriented" research. They see a need, instead, for "process-oriented" research, which can provide clues to the nature of the psi process. Presumably, this process can be investigated by the same methods which are used to study other psychological processes. In doing so, parapsychologists may commit methodological errors. But Rao sees these errors as occurring equally often in the other areas of behavioral science.

Those who do not accept the evidence for psi see the matter differently. They see some important differences between parapsychological research and research in other behavioral sciences. These differences result in their applying very high standards for what constitutes an adequate experiment in parapsychology. First, and most important, the claims of parapsychology are radical, in that they appear to require fundamental revisions or extensions of current physical theory. Second, the observed effects are generally small in size, and notoriously difficult to replicate. Third, the difficulty in replication appears to depend on the identity of the experimenter. This latter feature suggests normal explanations in terms of experimenter error. The combination of features provides legitimate grounds for skepticism, and for the application of strict standards of evidence, as I have advocated in this chapter.

J.G. Pratt, a longtime colleague of Rhine, called for stricter standards; he did see a need for "proof-oriented" research (which does not preclude study of the "psi process"). Without higher standards, Pratt perceived a danger to the field—that "parapsychologists will come to believe that our field is vigorously healthy while orthodox scientists will come to regard it as a cult" (Pratt, 1979, p. 24). This is a genuine risk.

References

Alcock, J.E. *Parapsychology: Science or magic? A psychological perspective.* Elmsford, N.Y.: Pergamon Press, 1981.

Barber, T.X. *Pitfalls in human research: Ten pivotal points.* Elmsford, N.Y.: Pergamon Press, 1976.

Beloff, J. Reviews of Uri Geller: *My story*, by Uri Geller; *Uri*, by A. Puharich; and *Super-minds, an enquiry into the paranormal*, by J. Taylor. *Journal of Parapsychology*, 1975, 39, 242-250.

Bevan, J.M. The relationship of attitude to success in ESP scoring. *Journal of Parapsychology*, 1947, 11, 296-309.

Bhadra, B.H. The relationship of test scores to belief in ESP. *Journal of Parapsychology*, 1966, 30, 1-17.

Bierman, D.J., & Houtkooper, J.M. Exploratory PK tests with a programmable high-speed random number generator. *European Journal of Parapsychology*, 1975-1977, 1(1), 3-14.

Bierman, D.J., & Houtkooper, J.M. The potential observer effect or the mystery of irreproducibility. *European Journal of Parapsychology*, 1981, 3(4), 345-372.

Boring, E.G. The present status of parapsychology. *American Scientist*, 1955, 43, 108-116.

Braud, L.W., & Braud, W.G. Further studies of relaxation as a psi-conducive state. *Journal of the American Society for Psychical Research*, 1974, 68, 229-245.

Braud, L.W., & Braud, W.G. Clairvoyance tests following exposure to a psi-conducive tape recording. *Journal of Research in Psi Phenomena*, 1977, 2(l), 10-2 1.

Braud, W.G., & Braud, L.W. Preliminary explorations of psi-conducive states: Progressive muscular relaxation. *Journal of the American Society for Psychical Research*, 1973, 67, 26-46.

Braud, W.G., & Mellen, R.R. A preliminary investigation of clairvoyance during hypnotic age regression. *European Journal of Parapsychology*, 1979, 2(4), 371-378.

Braud, W.G., & Wood, R. The influence of immediate feedback on free-response GESP performance during ganzfeld stimulation. *Journal of the American Society for PsychicalResearch*, 1977, 71, 409-427.

Braud, W.G., Wood, R., & Braud, L.W. Free-response GESP performance during an experimental hypnogogic state induced by visual and acoustic ganzfeld techniques: A replication and extension. *Journal of the American Society for Psychical Research*, 1975, 69, 105-114.

Bridgman, P.W. Probability, logic, and ESP. *Science*, 1956, 123, 15-17.

Broad, W.J. Magician's effort to foil scientists raises questions. *New York Times*, February 15,1983,19.

Broad, W.J., & Wade, N. *Betrayers of the truth.* New York: Simon & Schuster, 1982.

Broughton, R.S. An experiment with the head of Jut. *European Journal of Parapsychology*, 1979, 2, 337-357.

Broughton, R.S. Computer methodology: Total control with a human face. *Parapsychology Review*, 1982, 13(2), 1-6.

Brown, G.S. *Probability and scientific inference*. London: Longmans, 1957.

Burdick, D.S., & Kelly, E.F. Statistical methods in parapsychological research. In B.B. Wolman (Ed.), *Handbook of parapsychology*. New York: Van Nostrand Reinhold, 1977.

Calkins, J.L. Comments by James Calkins. (Commentary on the paper by Professor Karnes et al.) *Zetetic Scholar*, 1980, No. 6, 77-81.

Carpenter, J.C. The differential effect and hidden target differences consisting of erotic and neutral stimuli. *Journal of the American Society for Psychical Research*, 1971, 65, 204-214.

Casler, L. The improvement of clairvoyance scores by means of hypnotic suggestion. *Journal of Parapsychology*, 1962, 26, 77-87.

Casler, L. The effects of hypnosis on GESP. *Journal of Parapsychology*, 1964, 28, 126-134.

Casler, L. Hypnotic maximization of ESP motivation. *Journal of Parapsychology*, 1976, 40, 187-193.

Casper, G.W. Effect of receiver's attitude toward sender in ESP tests. *Journal of Parapsychology*, 1952, 16, 212-218.

Child, I.L., & Levi, A. Psi-missing in free-response settings. *Journal of the American Society for Psychical Research*, 1979, 73, 273-289.

Child, I.L., & Levi, A. The use of judges' ratings to test hypotheses about psi processes. *Journal of the American Society for Psychical Research*, 1980, 74, 171-181.

Child, I.L., & Levi, A. Psi-missing again in ganzfeld sessions? In W.G. Roll & J. Beloff (Eds.), *Research in parapsychology 1980*. Metuchen, N.J.: Scarecrow Press, 1981, 85-86.

Cochran, W.G., & Cox, G.M. *Experimental designs* (2nd ed.). New York: Wiley, 1957.

Crumbaugh, J.C. A scientific critique of parapsychology. *International Journal of Neuropsychiatry*, 1966, 2, 523-531.

Dale, L.A. Note on an attempt to repeat the Woodworth results. *Journal of the American Society for Psychical Research*, 1943, 37, 134-137.

Davis, J.W. A developmental program for the computer-based extension of parapsychological research and methodology. *Journal of Parapsychology*, 1974, 38, 69-84.

Davis, J.W. The stacking effect: Its practical significance in parapsychology. *Journal of Parapsychology*, 1978, 42, 67.

Davis, J.W., & Akers, C. Randomization and tests for randomness. *Journal of Parapsychology*, 1974, 38, 393-407.

Diaconis, P. Statistical problems in ESP research. *Science*, 1978, 201, 131-136.

Dingwall, E.J. Letter to the editor. *Journal of the Society for Psychical Research*, 1937, 13, 140-141.

Eisenbud, J. Review of The Geller papers, edited by C. Panati. *Journal of Parapsychology*,1976, 40, 321-325.

Estabrooks, G.H. *Spiritism*. New York: Dutton, 1947.

Fahler, J., & Cadoret, R.J. ESP card tests of college students with and without hypnosis. *Journal of Parapsychology*, 1958, 22, 125 -136.

Feller, W.K. Statistical aspects of ESP. *Journal of Parapsychology*, 1940, 4, 271-298.

Frazier, K. (Ed.). *Paranormal borderlines of science*. Buffalo: Prometheus Books, 1981.

Gardner, M. Mathematical games: Concerning an effort to demonstrate extrasensory perception by machine. *Scientific American*, October 1975, 238, 114-118.

Gardner, M. Review of *Learning to use extrasensory perception* by C. Tart. *New York Review of Books*, July 14, 1977, 37.

Gardner, M. *Science: Good, bad, and bogus*. Buffalo: Prometheus Books, 1981.

Gardner, M. How not to test a psychic: The great SRI die mystery. *Skeptical Inquirer*, 1982-1983, 7(2), 33-39.

Gatlin, L.L. Meaningful information creation: An alternative interpretation of the psi phenomenon. *Journal of the American Society for Psychical Research*, 1977, 71, 1-18.

Gatlin, L.L. Comments on the critical exchange between Drs. Stanford and Tart. *Journal of the American Society for Psychical Research*, 1978, 72, 77-81. (a)

Gatlin, L.L. Dr. Gatlin's reply to Dr. Tart. *Journal of the American Society for Psychical Research*, 1978, 72, 294-296. (b)

Gatlin, L.L. A new measure of bias in finite sequences with applications to ESP data. *Journal of the American Society, for Psychical Research*, 1979, 73, 29-43.

Girden, E. A review of psychokinesis (PK). *Psychological Bulletin*, 1962, 59, 353-388.

Girden, E. Parapsychology. In E.C. Carterette & M.P. Friedman (Eds.), *Handbook of perception* (Vol. 10). New York: Academic Press, 1978.

Goldman, A., Stein, S., & Weiner, H. Letter to the editor. *New York Review of Books*, July 14, 1977, 37.

Goodfellow, L.D. A psychological interpretation of the results of the Zenith Radio experiments. *Journal of Experimental Psychology*, 1938, 23, 601-632.

Greenwood, J.A. An empirical investigation of some sampling problems. *Journal of Parapsychology*, 1938, 2, 222-230.

Grela, J.J. Effect on ESP scoring of hypnotically induced attitudes. *Journal of Parapsychology*, 1945, 9, 194-202.

Greville, T.N.E. On multiple matching with one variable deck. *Annals of Mathematical Statistics*, 1944, 15, 432-434.

Hansel, C.E.M. *ESP: A scientific evaluation*. New York: Scribner's, 1966.

Hansel, C.E.M. *ESP and parapsychology: A critical re-evaluation*. Buffalo: Prometheus Books, 1980.

Hansel, C.E.M. A critical analysis of H. Schmidt's psychokinesis experiments. *Skeptical Inquirer*, 1981, 5, 26-33.

Haraldsson, E. ESP and the defense mechanism test (DMT): A further validation. *European Journal of Parapsychology*, 1978, 2, 104-114.

Hardy, A., Harvie, R., & Koestler, A. *The challenge of chance*. London: Hutchinson, 1973.

Herbert, C.V.C. Experiment in extra-sensory perception: 1. A note on types of Zener cards used at Duke University. *Journal of the Society for Psychical Research*, 1938, 32, 215-218.

Honorton, C. Separation of high- and low-scoring ESP subjects through hypnotic preparation. *Journal of Parapsychology*, 1964, 28, 251-257.

Honorton, C. A further separation of high- and low-scoring ESP subjects through hypnotic preparation. *Journal of Parapsychology*, 1966, 30, 172-183,

Honorton, C. Significant factors in hypnotically-induced clairvoyant dreams. *Journal of the American Society for Psychical Research*, 1972, 66, 86-102.

Honorton, C. Error some place! *Journal of Communication*, 1975, 2.5, 103-116. (a)

Honorton, C. Psi and internal attention states. In B.B. Wolman (Ed.), *Handbook of parapsychology*. New York: Van Nostrand Reinhold, 1977.

Honorton, C. *Replicability, experimenter influence, and parapsychology: An empirical context for the study of mind*. Paper presented at the meeting of the American Association for the Advancement of Science, Washington, D.C., 1978. (b)

Honorton, C. Methodological issues in free-response psi experiments. *Journal of the American Society for Psychical Research*, 1979, 73, 381-394.

Honorton, C. Beyond the reach of sense: Some comments on C.E.M. Hansel's *ESP and parapsychology: A critical re-evaluation. Journal of the American Society for Psychical Research*, 1981, 75, 155-166.

Honorton, C. Response to Hyman's critique of psi ganzfeld studies. In W.G. Roll, J. Beloff, & R.A. White (Eds.), *Research in parapsychology 1982*. Metuchen, N.J.: Scarecrow Press, 1983, 23-26.

Honorton, C., & Harper, S. Psi-mediated imagery and ideation in an experimental procedure for regulating perceptual input. *Journal of the American Society for Psychical Research*, 1974, 68, 156-168.

Honorton, C., & Stump, J. A preliminary study of hypnotically-induced clairvoyant dreams. *Journal of the American Society for Psychical Research*, 1969, 63, 175-184.

Humphrey, B.M. Further work of Doctor Stuart on interest ratings and ESP. *Journal of Parapsychology*, 1949, 13, 151-165.

Humphreys, W.C. *Anomalies and scientific theories*. San Francisco: Freeman, Cooper, 1968.

Hyman, R. Review of *The Geller papers*, edited by C. Panati. *The Zetetic*, 1(1), 1976, 73-80.

Hyman, R. The case against parapsychology. *The Humanist*, November/December 1977, 47-49. (a)

Hyman, R. Psychics and scientists: "Mind-Reach" and remote viewing. *The Humanist*, May/June 1977, 16-20. (b)

Hyman, R. Further comments on Schmidt's PK experiments: Alternative explanations are abundant. *Skeptical Inquirer*, 1981, 5, 34-40. (a)

Hyman, R. Scientists and psychics. In G.O. Abell & B. Singer (Eds.), *Science and the paranormal*. New York: Scribner's, 1981. (b)

Hyman, R. Does the ganzfeld experiment answer the critics' objections? In W.G. Roll, J. Beloff, & R.A. White (Eds.), *Research in parapsychology 1982*. Metuchen, N.J.: Scarecrow Press, 1983, 21-23.

Johnson, M. Models of control and control of bias. *European Journal of Parapsychology*, 1975, 1(1), 36-44.

Johnson, M., & Kanthamani, B.K. The Defense Mechanism Test as a predictor of ESP scoring direction. *Journal of Parapsychology*, 1967, 31, 99-110.

Kanthamani, B.K. *The ESP subject: An inquiry into the personality patterns of psi hitters and missers*. Unpublished doctoral dissertation, Andhra University, 1969.

Kanthamani, B.K., & Rao, K.R. Personality characteristics of ESP subjects: 1. Primary personality characteristics and ESP. *Journal of Parapsychology*, 1971, 35, 189-207.

Kellogg, C.E. New evidence (?) for extra-sensory perception. *Scientific Monthly*, 1937, 45, 331-341.

Kelly, E.F. Review of *The challenge of chance* by A. Hardy, R. Harvie, & A. Koestler. *Parapsychology Review*, 1974, 5(1), 16-19.

Kelly, E.F. Reply to Persi Diaconis. Zetetic Scholar, 1979, No. 5, 20-28.

Kennedy, J.E. Methodological problems in free-response ESP experiments. *Journal of the American Society for Psychical Research*, 1979, 73, 1-15. (a)

Kennedy, J.E. More on methodological issues in free-response psi experiments. *Journal of the American Society for Psychical Research*, 1979, 73, 395-401. (b)

Kennedy, J.E. Ambiguous data result in ambiguous conclusions: A reply to Charles T. Tart. *Journal of the American Society for Psychical Research*, 1980, 74, 349-356. (a)

Kennedy, J.E. Learning to use ESP: Do the calls match the targets or do the targets match the calls? *Journal of the American Society for Psychical Research*, 1980, 74, 191-209. (b)

Kennedy, J.E., & Taddonio, J.L. Experimenter effects in parapsychological research. *Journal of Parapsychology*, 1976, 40, 1-33.

Kennedy, J.L. The visual cues from the backs of ESP cards. *Journal of Psychology*, 1938, 6, 149-153.

Kesner, J., & Morris, R.L. A precognition test using guided imagery. In W.G. Roll (Ed.),*Research in parapsychology 1977*. Metuchen, N.J.: Scarecrow Press, 1978, 48-52.

Krippner, S. Experimentally-induced telepathic effects in hypnosis and nonhypnosis groups. *Journal of the American Society for Psychical Research*, 1968, 62, 387-398.

Lemmon, V.W. Extra-sensory perception. *Journal of Psychology*, 1937, 4, 227-238.

Leuba, L. An experiment to test the role of chance in ESP research. *Journal of Parapsychology*, 1938, 2, 387-398.

Loftus, E.F. *Eyewitness testimony*. Cambridge, Mass.: Harvard University Press, 1979.

Marks, D. On the review of *The psychology of the psychic*: A reply to Dr. Morris. *Journal of the American Society for Psychical Research*, 1981, 75, 197-203. (a)

Marks, D. Sensory cues invalidate remote viewing experiments. *Nature*, 1981, 292, 177. (b)

Marks, D. Remote viewing revisited. *Skeptical Inquirer*, 1982, 6, 18-29.

Marks, D., & Kammann, R. The nonpsychic powers of Uri Geller. *The Zetetic*, 1977, 1(2), 9-17.

Marks, D., & Kammann, R. Information transmission in remote viewing experiments. *Nature*, 1978, 274, 680-681.

Marks, D., & Kammann, R. *The psychology of the psychic*. Buffalo: Prometheus Books, 1980.

Markwick, B. The Soal-Goldney experiments with Basil Shackleton: New evidence of data manipulation. *Proceedings of the Society for Psychical Research*, 1978, 56, 250-277.

Mauskopf, S.H., & McVaugh, M.R. The controversy over statistics in parapsychology: 1934-1938. In S.H. Mauskopf (Ed.), *The reception of unconventional science (AAAS Selected Symposium 25)*. Boulder, Colo.: Westview, 1979.

Mauskopf, S.H., & McVaugh, M.R. *The elusive science: Origins of experimental psychical research*. Baltimore: Johns Hopkins University Press, 1980.

May, E.C., Humphrey, B.S., & Hubbard, G.S. *Electronic system perturbation techniques*. Menlo Park, Calif.: SRI International, 1980.

McBain, W., Fox, W., Kimura, S., Nakanishi, M., & Tirado, J. Quasi-sensory communication: An investigation using semantic matching and accentuated affect. *Journal of Personality and Social Psychology*, 1970, 14, 281-291.

McConnell, R.A. Comments on C. Scott's review of Extrasensory, perception and personality patterns. *Journal of the Society for Psychical Research*, 1959, 40, 79.

McConnell, R.A. A parapsychological dialogue. *Journal of the American Society for Psychical Research*, 1977, 71, 429-435.

McConnell, R.A., & Clark, T.K. Training, belief, and mental conflict within the Parapsychological Association. *Journal of Parapsychology*, 1980, 44, 245-268.

McConnell, R.A., Snowdon, R.J., & Powell, K.F. Wishing with dice. *Journal of Experimental Psychology*, 1955, 50, 269-275.

Medhurst, R.G., & Scott, C. A re-examination of C.E.M. Hansel's criticism of the Pratt-Woodruff experiment. *Journal of Parapsychology*, 1974, 38, 163-184.

Morris, R.L. An exact method for evaluating preferentially matched free-response material. *Journal of the American Society for Psychical Research*, 1972, 66, 401-407.

Morris, R.L. Some comments on the assessment of parapsychological studies: A review of *The psychology of the psychic* by David Marks and Richard Kammann. *Journal of the American Society for Psychical Research*, 1980, 74, 425-443. (b)

Morris, R.L. Dr. Morris replies to Dr. Marks. *Journal of the American Society for Psychical Research*, 1981, 75, 203-207.

Morris, R.L. Review of *Parapsychology: Science or magic? A psychological perspective*, by J.E. Alcock. *Journal of the American Society for Psychical Research*, 1982, 76, 177-186.

Moss, S., & Butler, D.C. The scientific credibility of ESP. *Perceptual and Motor Skills*, 1978, 46, 1063-1074.

Moss, T., & Gengerelli, J.A. ESP effects generated by affective states. *Journal of Parapsychology*, 1968, 32, 90-100.

Moss, T., Paulson, M., Chang, A., & Levitt, M. Hypnosis and ESP: A controlled experiment. *Journal of Clinical Hypnosis*, 1970, 13, 46-56.

Murphy, G. On limits of recording errors. In The ESP symposium at the A.P.A. *Journal of Parapsychology*, 1938, 2, 262-266.

Murphy, G. Report on a paper by Edward Girden on psychokinesis. *Psychological Bulletin*, 1962, 59, 520-528.

Musso, J.R. ESP experiments with primary school children. *Journal of Parapsychology*, 1965, 29, 115-121.

Nash, C.S., & Nash, C.B. An ESP experiment with targets that differ in degree of similarity. *Journal of the American Society for Psychical Research*, 1961, 55, 73-76.

Nicol, J.F. The statistical controversy in quantitative research. *International Journal of Parapsychology*, 1959, 1, 47-63.

Nicol, J.F. The experimenter's responsibility. In M. Ebon (Ed.), *The Satan trap*. Garden City, N.Y.: Doubleday, 1976.

Nicol, J.F., & Humphrey, B.M. The exploration of ESP and human personality. *Journal of the American Society for Psychical Research*, 1953, 47, 133-178.

Palmer, J. Scoring in ESP tests as a function of belief in ESP: Part 1. The sheep-goat effect. *Journal of the American Society for Psychical Research*, 1971, 65, 373-408.

Palmer, J. Attitudes and personality traits in experimental ESP research. In B.B. Wolman (Ed.), *Handbook of parapsychology*. New York: Van Nostrand Reinhold, 1977.

Palmer, J. Extrasensory perception: Research findings. In S. Krippner (Ed.), *Advances in parapsychological research. Vol. 2: Extrasensory perception*. New York: Plenum, 1978.

Palmer, J. *Methodological objections to the case for psi: Are formal control conditions necessary for the demonstration of psi?* Paper presented at the meeting of the American Psychological Association, Los Angeles, Calif., 1981.

Palmer, J. ESP research findings: 1976-1978. In S. Krippner (Ed.), *Advances in parapsychological research. Vol. 3*. New York: Plenum, 1982.

Palmer, J., & Lieberman, R. The influence of psychological set on ESP and out-of-body experiences. *Journal of the American Society for Psychical Research*, 1975, 69, 193-213.

Palmer, J., & Vassar, C. ESP and out-of-body experiences: An exploratory study. *Journal of the American Society for Psychical Research*, 1974, 68, 257-280.

Panati, C. (Ed.) *The Geller papers*. Boston: Houghton Mifflin, 1976.

Parker, A. A holistic methodology in psi research. *Parapsychology Review*, 1978, 9(2), 1-6.

Parker, A., & Beloff, J. Hypnotically-induced clairvoyant dreams: A partial replication and attempted confirmation. *Journal of the American Society for Psychical Research*, 1970, 64, 432-442.

Phillips, P.R., & Shafer, M. Exploratory research with two new psychic metal-benders. In W.G. Roll, R.L. Morris, & R.A. White (Eds.), *Research in parapsychology 1981*. Metuchen, N.J.: Scarecrow Press, 1982, 144-146.

Pope, D., & Pratt, J.G. The ESP controversy. *Journal of Parapsychology*, 1942, 6, 174-189.

Pratt, J.G. Is ESP only a misnomer for response sequences chosen to match inferred target sequences? A reply to Lila L. Gatlin. In W.G. Roll (Ed.), *Research in parapsychology 1978*. Metuchen, N.J.: Scarecrow Press, 1979, 129-131. (a)

Pratt, J.G. Parapsychology, normal science, and paradigm change. *Journal of the American Society for Psychical Research*, 1979, 73, 17-28. (b)

Pratt, J.G., Rhine, J.B., Smith, B.M., Stuart, C.E., & Greenwood, J.A. *Extra-sensory perception after sixty years*. Boston: Bruce Humphries, 1966. (Originally published, 1940.)

Pratt, J.G., & Woodruff, J.L. Size of stimulus symbols in extra-sensory perception. *Journal of Parapsychology*, 1939, 3, 121-158.

Price, G.R. Science and the supernatural. *Science*, 1955, 122, 359-367.

Price, G.R. Apology to Rhine and Soal. *Letter to Science*, 1972, 175, 359.

Puthoff, H.E., & Targ, R. The record: Eight days with Uri Geller. In C. Panati (Ed.), *The Geller papers*. Boston: Houghton Mifflin, 1976.

Rand Corporation. A million random digits. Glencoe, Ill.: Free Press, 1955.

Randi, J. *The magic of Uri Geller*. New York: Ballantine, 1975.

Randi, J. *Flim-flam! The truth about unicorns, parapsychology and other delusions*. New York: Lippincott & Crowell, 1980.

Rao, K.R. Psi: Its place in nature. *Journal of Parapsychology*, 1978, 42, 276-303.

Rao, K.R. On "The scientific credibility of ESP." *Perceptual and Motor Skills*, 1979, 49, 415-429.

Rao, K.R. Comments on "Seven evidential experiments." *Zetetic Scholar*, 1980, No. 6, 107-109.

Rao, K.R. Letter to the editor. *Journal of the Society for Psychical Research*, 1981, 51, 191-193.

Rao, K.R., Dukhan, H., & Rao, P.V.K. Yogic meditation and psi scoring in forced-choice and free-response tests. *Journal of Indian Psychology*, 1978, 1, 1-16.

Rhine, J.B. *Extra-sensory perception*. Boston: Bruce Humphries, 1964. (Originally published, 1934.)

Rhine, J.B. Comments: A new case of experimenter unreliability. *Journal of Parapsychology*, 1974, 38, 215-225. (a)

Rhine, J.B. Comments: Security versus deception in parapsychology. *Journal of Parapsychology*, 1974, 38, 99-121. (b)

Rhine, J.B. Comments: Second report on a case of experimenter fraud. *Journal of Parapsychology*, 1975, 39, 306-325.

Rhine, J.B., & Humphrey, B.M. The PK effect: Special evidence from hit patterns. 1. Quarter distribution of the page. *Journal of Parapsychology*, 1944, 8, 18-60.

Rhine, J.B., & Pratt, J.G. A review of the Pearce-Pratt distance series of ESP tests. *Journal of Parapsychology*, 1954, 18, 165-177.

Rhine, J.B., & Pratt, J.G. *Parapsychology: Frontier science of the mind* (2nd rev. ed.) Springfield, Ill.: Thomas, 1962.

Rogo, D.S. Letter to the editor. *Journal of Parapsychology*, 1979, 43, 266. (a)

Rosenthal, R. *Experimenter effects in behavioral research* (rev. ed.). New York: Irvington, 1976.

Rosenthal, R. How often are our numbers wrong? *American Psychologist*, 1978, 33, 1005-1008.

Rosenthal, R., & Rubin, D.B. Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences*, 1978, 3, 377-415.

Roth, J.A. Hired hand research. *American Sociologist*, 1966, 1, 190-196.

Ryzl, M. Precognition scoring and attitude toward ESP. *Journal of Parapsychology*, 1968, 32, 1-8.

Sargent, C.L. Hypnosis as a psi-conducive state: A controlled replication study. *Journal of Parapsychology*, 1978, 42, 257-275. (a)

Sargent, C.L. Comments on "Effects of associations and feedback on psi in the ganzfeld." Letter to the editor. *Journal of the American Society for Psychical Research*, 1980, 74, 265-267. (a)

Sargent, C.L. *Exploring psi in the ganzfeld*. New York: Parapsychology Foundation, 1980. (b)

Sargent, C.L., Bartlett, H.J., & Moss, S.P. Response structure and temporal incline in ganzfeld free-response GESP testing. *Journal of Parapsychology*, 1982, 46, 85-110.

Sargent, C.L., & Harley, T.A. Three studies using a psi-predictive trait variable questionnaire. *Journal of Parapsychology*, 1981, 4.5, 199-214.

Schmeidler, G.R. Predicting good and bad scores in a clairvoyance experiment: A final report. *Journal of the American Society for Psychical Research*, 1943, 37, 210-221.

Schmeidler, G.R. Separating the sheep from the goats. *Journal of the American Society for Psychical Research*, 1945, 39, 47-49.

Schmeidler, G.R. Letter to the editor. *Journal of the Society for Psychical Research*, 1951, 36, 410-412.

Schmeidler, G.R. Comments and additional data on sheep-goat classification and targets. *Journal of the Society for Psychical Research*, 1959, 40, 63-72.

Schmeidler, G.R. High ESP scores after a swami's brief instruction in meditation and breathing. *Journal of the American Society for Psychical Research*, 1970, 64, 100-113.

Schmeidler, G.R. Mood and attitude on a pretest as predictors of retest ESP performance. *Journal of the American Society for Psychical Research*, 1971, 65, 324-335.

Schmeidler, G.R. Methods for controlled research on ESP and PK. In B.B. Wolman (Ed.), *Handbook of parapsychology*. New York: Van Nostrand Reinhold, 1977. (a)

Schmeidler, G.R., & McConnell, R.A. *ESP and personality patterns*. New Haven: Yale University Press, 1958.

Schmeidler, G.R., & Murphy, G. The influence of belief and disbelief in ESP upon ESP scoring level. *Journal of Experimental Psychology*, 1946, 36, 271-276.

Schmidt, H. Clairvoyance tests with a machine. *Journal of Parapsychology*, 1969, 33, 300-307. (a)

Schmidt, H. Precognition of a quantum process. *Journal of Parapsychology*, 1969, 33, 99-108. (b)

Schmidt, H. A PK test with electronic equipment. *Journal of Parapsychology*, 1970, 34, 175-181. (a)

Schmidt, H. A quantum mechanical random number generator for psi tests. *Journal of Parapsychology*, 1970, 34, 219-224. (b)

Schmidt, H. A program for channeling psi data into the laboratory and onto the critic's desk. In W.G. Roll (Ed.), *Research in parapsychology 1979*. Metuchen, N.J.: Scarecrow Press, 1980, 66-69.

Schmidt, H., Morris, R.L., & Rudolph, L. *Channeling evidence for a psychokinetic effect to critical observers*. Paper presented at the meeting of the Parapsychological Association; Madison, N.J., 1983.

Schmitt, M., & Stanford, R.G. Free-response ESP during ganzfeld stimulation: The possible influence of menstrual cycle phase. *Journal of the American Society for Psychical Research*, 1978, 72, 177-182.

Schouten, S.A., & Kelly, E.F. On the experiment of Brugmans, Heymans, and Weinberg. *European Journal of Parapsychology*, 1978, 2(3), 247-290.

Scott, C. Letter to the editor. *Journal of the Society for Psychical Research*, 1951, 36, 409-410. (a)

Scott, C. Letter to the editor. *Journal of the Society for Psychical Research*, 1951, 36, 452. (b)

Scott, C. Spencer Brown and probability: A critique. *Journal of the Society for Psychical Research*, 1958, 39, 217-234.

Scott, C. *Review of Extrasensory perception and personality patterns* by G.R. Schmeidler & R.A. McConnell. *Journal of the Society for Psychical Research*, 1959, 40, 73-79.

Scott, C. On the evaluation of verbal material in parapsychology: A discussion of Dr. Pratt's monograph. *Journal of the Society for Psychical Research*, 1972, 46, 79-90.

Scott, C. Christopher Scott replies to John Beloff and Sybo Schouten (ZS #6). *Zetetic Scholar*, 1980, No. 7, 140-143.

Scott, C., & Haskell, P. Fresh light on the Shackleton experiments? *Proceedings of the Society for Psychical Research*, 1974, 56, 43-72.

Sells, S.B. Letter to the editor: Review of "Terminal salience in ESP performance," by J.B. Rhine. *Journal of Parapsychology*, 1941, 5, 250-254.

Shields, E. Comparison of children's guessing ability (ESP) with personality characteristics. *Journal of Parapsychology*, 1962, 26, 200-210.

Shrager, E.F. The effect of sender-receiver relationship and associated personality variables on ESP scoring in young children. *Journal of the American Society for Psychical Research*, 1978, 72, 35-47.

Sondow, N. Effects of associations and feedback on psi in the ganzfeld: Is there more than meets the judge's eye? *Journal of the American Society for Psychical Research*, 1979, 73, 123-150.

Sondow, N. Letter to the editor. *Journal of the American Society for Psychical Research*, 1980, 74, 267-272.

Sondow, N., Braud, L., & Barker, P. Target qualities and affect measures in an exploratory psi ganzfeld. In W.G. Roll, R.L. Morris, & R.A. White (Eds.), *Research in parapsychology 1981*. Metuchen, N.J.: Scarecrow Press, 1982, 82-85.

Stanford, R.G. Experimental psychokinesis. In B.B. Wolman (Ed.), *Handbook of parapsychology*. New York: Van Nostrand Reinhold, 1977. (b)

Stanford, R.G. The question is: Good experimentation or not? A reply to Dr. C.T. Tart. *Journal of the American Society for Psychical Research*, 1977, 71, 191-200. (c)

Stanford, R.G. Review of The application of learning theory to ESP performance by C.T. Tart. *Journal of the American Society for Psychical Research*, 1977, 71, 55-80. (d)

Stanford, R.G. The influence of auditory ganzfeld characteristics upon free-response ESP performance. *Journal of the American Society for Psychical Research*, 1979, 73, 253-272.

Stanford, R.G. Review of Exploring psi in the ganzfeld by Carl L. Sargent. *Journal of theAmerican Society for Psychical Research*, 1981, 75, 348-357. (d)

Stanford, R.G., & Mayer, B. Relaxation as a psi-conducive state: A replication and exploration of parameters. *Journal of the American Society for Psychical Research*, 1974, 68, 182-191.

Stokes, D. Review of Mind-reach: Scientists look at psychic ability, by R. Targ & H. Puthoff. *Journal of the American Society for Psychical Research*, 1977, 71, 437-442.

Stokes, D. Review of *Research in Parapsychology 1976*, edited by J.D. Morris, W.G. Roll, & R.L. Morris. *Journal of Parapsychology*, 1978, 42, 70-76.

Stokes, D. Review of *Mind at large*, edited by C.T. Tart, H.E. Puthoff, & R. Targ. *Journal of the American Society for Psychical Research*, 1980, 74, 445-454.

Targ, R., & Puthoff, H. Information transmission under conditions of sensory shielding. *Nature*, 1974, 252, 602-607.

Targ, R., & Puthoff, H. Letter to the editor. *Scientific American*, January 1976, 239, 6.

Targ, R., & Puthoff, H. *Mind-reach: Scientists look at psychic ability*. New York: Delacorte, 1977.

Targ, R., Cole, P., & Puthoff, H. *Development of techniques to enhance man/machine communication*. Menlo Park, Calif.: Stanford Research Institute, 1974.

Tart, C.T. *The application of learning theory to ESP performance*. New York: Parapsychology Foundation, 1975.

Tart, C.T. Letter to the editor. *New York Review of Books*, October 13, 1977. (a)

Tart, C.T. Toward humanistic experimentation in parapsychology: A reply to Dr. Stanford's review. *Journal of the American Society for Psychical Research*, 1977, 71, 81-101. (b)

Tart, C.T. Letter to the editor: Dr. Tart's reply to Dr. Gatlin. *Journal of the American Society for Psychical Research*, 1978, 72, 81-87.

Tart, C.T. Could mathematical inference strategies have inflated scoring in the first feedback training study? In W.G. Roll (Ed.), *Research in parapsychology 1978*. Metuchen, N.J.: Scarecrow Press, 1979, 128-129. (a)

Tart, C.T. Randomicity, predictability, and mathematical inference strategies in ESP feedback experiments: Discussion of Dr. Gatlin's paper. *Journal of the American Society for Psychical Research*, 1979, 73, 44-60. (b)

Tart, C.T. Are we interested in making ESP function strongly and reliably? *Journal of the American Society for Psychical Research*, 1980, 74, 210-222.

Tart, C.T. Are prepared random sequences and real-time random generators interchangeable? In W.G. Roll & J. Beloff (Eds.), *Research in parapsychology 1980*. Metuchen, N.J.: Scarecrow Press, 1981, 43-45. (a)

Tart, C.T. Letter to the editor. *New York Review of Books*, February 19, 1981. (b)

Tart, C.T., & Dronek, E. Trying to profit from nonrandomicity in ESP target sequences: Initial explorations with the Probabilistic Predictor Program. In W.G. Roll (Ed.), *Research in parapsychology 1979*. Metuchen, N.J.: Scarecrow Press, 1980, 86-88.

Tart, C.T., Palmer, J., & Redington, D.J. Effects of immediate feedback on ESP erformance: A second study. *Journal of the American Society for Psychical esearch*, 1979, 73, 151-165.

Tart, C.T., Puthoff, H.E., & Targ, R. Information transmission in remote viewing experiments, *Nature*, 1980, 284, 191.

Terry, J.C. A multiple session ganzfeld study. In J.D. Morris, W.G. Roll, & R.L. Morris (Eds.), *Research in parapsychology 1974*. Metuchen, N.J.: Scarecrow Press, 1975, 48-49.

Terry, J.C., & Honorton, C. Psi information retrieval in the ganzfeld: Two confirmatory studies. *Journal of the American Society for Psychical Research*, 1976, 70, 207-217.

Truzzi, M. A skeptical look at Paul Kurtz's analysis of the scientific status of parapsychology. *Journal of Parapsychology*, 1980, 44, 35-55.

Ullman, M., & Krippner, S. *Dream studies and telepathy: Experimental series VIII.* Preliminary report. Unpublished manuscript, July 1969.

Ullman, M., & Krippner, S. *Dream studies and telepathy*. New York: Parapsychology Foundation, 1970.

Ullman, M., Krippner, S., & Vaughan, A. *Dream telepathy*. New York: Macmillan, 1973.

Uphoff, H.F., & Kennedy, J.L. Experiments on the nature of extrasensory perception: III.The recording error criticism of extra-chance scores. *Journal of arapsychology*, 1939, 3, 226-245.

White, R.A. The influence of the experimenter motivation, attitudes, and methods of handling subjects in psi test results. In B.B. Wolman (Ed.), *Handbook of arapsychology*. New York: Van Nostrand Reinhold, 1977.

Wicklund, N. Parapsychological temptations: A discussion of decline effects, sex effects, and post-hoc effects. In J.D. Morris, W.G. Roll, and R.L. Morris (Eds.), *Research in parapsychology 1976*. Metuchen, N.J.: Scarecrow Press, 1977, 125-128.

Wilson, R. Deviations from randomness in ESP experiments. *International Journal of Parapsychology*, 1966, 8, 387-395.

Wilson, W.R. Do parapsychologists really believe in ESP? *Journal of Social Psychology*, 1964, 64, 379-389.

Windle, C. Test-retest effect on personality questionnaires. *Educational and Psychological Measurement*, 1954, 14, 617-633.

Wolman, B.B. (Ed.). *Handbook of parapsychology*. New York: Van Nostrand Reinhold, 1977.

Zusne, L., & Jones, W.H. *Anomalistic psychology: A study of extraordinary phenomena of behavior and experience*. Hillsdale, N.J.: Erlbaum, 1982,

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[15]

Psychology and Anomalous Observations

The Question of ESP in Dreams

Irvin L. Child *Yale University*

ABSTRACT: Books by psychologists purporting to offer critical reviews of research in parapsychology do not use the scientific standards of discourse prevalent in psychology. Experiments at Maimonides Medical Center on possible extrasensory perception (ESP) in dreams are used to illustrate this point. The experiments have received little or no mention in some reviews to which they are clearly pertinent. In others, they have been so severely distorted as to give an entirely erroneous impression of how they were conducted. Insofar as psychologists are guided by these reviews, they are prevented from gaining accurate information about research that, as surveys show, would be of wide interest to psychologists as well as to others.

In recent years, evidence has been accumulating for the occurrence of such anomalies as telepathy and psychokinesis, but the evidence is not totally convincing. The evidence has come largely from experiments by psychologists who have devoted their careers mainly to studying these anomalies, but members of other disciplines, including engineering and physics, have also taken part. Some psychologists not primarily concerned with parapsychology have taken time out from other professional concerns to explore such anomalies for themselves. Of these, some have joined in the experimentation (e.g., Crandall & Hite, 1983; Lowry, 1981; Radin, 1982). Some have critically reviewed portions of the evidence (e.g., Akers, 1984; Hyman, 1985). Some, doubting that the phenomena could be real, have explored nonrational processes that might encourage belief in their reality (e.g., Ayeroff & Abelson, 1976). Still others, considering the evidence substantial enough to justify a constructive theoretical effort, have struggled to relate the apparent anomalies to better established knowledge in a way that will render them less anomalous (e.g., Irwin, 1979) or not anomalous at all (e.g., Blackmore, 1984). These psychologists differ widely in their surmise about whether the apparent anomalies in question will eventually be judged real or illusory; but they appear to agree that the evidence to date warrants serious consideration.

Serious consideration of apparent anomalies seems an essential part of the procedures of science, regardless of whether it leads to an understanding of new discoveries or to an understanding of how persuasive illusions arise. Apparent anomalies—just like the more numerous observations that are not anomalous—can receive appropriate attention only as they become accurately known to the scientists to whose work they are relevant. Much parapsychological research is barred from being seriously considered because it is either neglected or misrepresented in writings by some psychologists—among them, some who have placed themselves in a prime position to mediate interaction between parapsychological research and the general body of psychological knowledge. In this article, I illustrate this important general point with a particular case, that of experimental research on possible ESP in dreams. It is a case of especially great interest but is not unrepresentative of how psychological publications have treated similar anomalies.

The Maimonides Research

The experimental evidence suggesting that dreams may actually be influenced by ESP comes almost entirely from a research program carried out at the Maimonides Medical Center in Brooklyn, New York. Among scientists active in parapsychology, this program is widely known and greatly respected. It has had a major indirect influence on the recent course of parapsychological research, although the great expense of dream-laboratory work has prevented it from being a direct model.

None of the Maimonides research was published in the journals that are the conventional media for psychology. (The only possible exception is that a summary of one study [Honorton, Krippner, & Ullman, 1972] appeared in convention proceedings of the American Psychological Association.) Much of it was published in the specialized journals of parapsychology. The rest was published in psychiatric or other medical journals, where it would not be noticed by many psychologists. Most of it was summarized in popularized form in a book (Ullman, Krippner, & Vaughan, 1973) in which two of the researchers were joined by a popular writer whose own writings are clearly not in the scientific tradition, and the book departs from the pattern of scientific reporting that characterizes the original research reports.

How, then, would this research come to the attention of psychologists, so that its findings or its errors might in time be evaluated for their significance to the body of systematic observations upon which psychology has been and will be built? The experiments at Maimonides were published between about 1966 and 1972. In the years since—now over a decade—five books have been published by academic psychologists that purport to offer a scholarly review and evaluation of parapsychological research. They vary in the extent to which they seem addressed to psychologists themselves or to their students, but they seem to be the principal route by which either present or future psychologists, unless they have an already established interest strong enough to lead them to search out the original publications, might become acquainted with the experiments on ESP in dreams. I propose to review how these five books have presented knowledge about the experiments. First, however, I must offer a summary of the experiments; without that, my review would make sense only to readers already well acquainted with them.

The experiments at Maimonides grew out of Montague Ullman's observations, in his psychiatric practice, of apparent telepathy underlying the content of some dreams reported by his patients—observations parallel to those reported by many other psychiatrists. He sought to determine whether this apparent phenomenon would appear in a sleep laboratory under controlled conditions that would seem to exclude interpretations other than that of ESP. He was joined in this research by psychologist Stanley Krippner, now at the Saybrook Institute in San Francisco, and a little later by Charles Honorton, now head of the Psychophysical Research Laboratories in Princeton, New Jersey. Encouraged by early findings but seeking to improve experimental controls and identify optimal conditions, these researchers, assisted by numerous helpers and consultants, tried out various modifications of procedure. No one simple description of procedure, therefore, can be accurate for all of the experiments. But the brief description that follows is not, I believe, misleading as an account of what was generally done.

The Experimental Procedure

A subject would come to the laboratory to spend the night there as would-be percipient in a study of possible telepathic influence on dreams. He or she met and talked with the person who was going to serve as agent (that is, the person who would try to send a telepathic message), as well as with the two experimenters taking part that night, and procedures were explained in detail unless the percipient was a repeater for whom that step was not necessary. When ready to go to bed, the percipient was wired up in the usual way for monitoring of brain waves and eye movements, and he or she had no further contact with the agent or agent's experimenter until after the session was completed. The experimenter in the next room monitored the percipient's sleep and at the beginning of each period of rapid eye movements (REM), when it was reasonably certain the sleeper would be dreaming, notified the agent by pressing a buzzer.

The agent was in a remote room in the building, provided with a target picture (and sometimes accessory material echoing the theme of the picture) randomly chosen from a pool of potential targets as the message to be concentrated on. The procedure for random choice of a target from the pool was designed to prevent anyone else from knowing the identity of the target. The agent did not open the packet containing the target until isolated for the night (except for the one-way buzzer communication). Whenever signaled that the percipient had entered a REM period, the agent was to concentrate on the target, with the aim of communicating it telepathically to the percipient and thus influencing the dream the percipient was having. The percipient was oriented toward trying to receive this message. But of course if clairvoyance and telepathy are both possible, the percipient might have used the former—that is, might have been picking up information directly from the target picture, without the mediation of the agent's thoughts or efforts. For this reason, the term *general extrasensory perception* (*GESP*) would be used today, though the researchers more often used the term *telepathy.*

Toward the approximate end of each REM period, the percipient was awakened (by intercom) by the monitoring experimenter and described any dream just experienced (with prodding and questioning, if necessary, though the percipient of course knew in advance what to do on each awakening). At the end of the night's sleep, the percipient was interviewed and was asked for impressions about what the target might have been. (The interview was of course double-blind; neither percipient nor interviewer knew the identity of the target.) The dream descriptions and morning impressions and associations were recorded and later transcribed.

The original research reports and the popular book both present a number of very striking similarities between passages in the dream transcripts and the picture that happened to be the night's target. These similarities merit attention, yet they should in themselves yield no sense of conviction. Perhaps any transcript of a night's dreaming contains passages of striking similarity to any picture to which they might be compared. The Maimonides research, however, consisted of carefully planned experiments designed

Requests for reprints should be sent to Irvin L. Child at the Department of Psychology, Yale University, P.O. Box 11A, New Haven, Connecticut 06520-7447.

to permit evaluation of this hypothesis of random similarity, and I must now turn to that aspect.

Results

To evaluate the chance hypothesis, the researchers obtained judgments of similarity between the dream content and the actual target for the night, and at the same time obtained judgments of similarity between the dream content and each of the other potential targets in the pool from which the target had been selected at random. The person judging, of course, had no information about which picture had been randomly selected as target; the entire pool (in duplicate) was presented together, with no clue as to which picture had been the target and which ones had not. That is, in the experimental condition a picture was randomly selected from a pool and concentrated on by the agent, and in the control condition a picture was left behind in the pool. Any consistent difference between target and nontarget in similarity to dream content, exceeding what could reasonably be ascribed to chance, was considered an apparent anomaly.

The data available for the largest number of sessions came from judgments made by judges who had no contact with the experiment except to receive (by mail, generally) the material necessary for judging (transcripts of dreams and interview and a copy of the target pool). For many sessions, judgments were also available from the dreamer; he or she, of course, made judgments only after completing participation in the experiment as dreamer (except in some series where a separate target pool was used for each night and the dreamer's judgments could be made at the end of the session). For many sessions, judgments were made for the dream transcripts alone and for the total transcript including the morning interview; for consistency I have used the latter, because it involved judges who had more nearly the same information as the subjects.

The only form in which the data are available for all series of sessions is a count of hits and misses. If the actual target was ranked in the upper half of the target pool, for similarity to the dreams and interview, the outcome was considered a hit. If the actual target was ranked in the lower half of the pool, the outcome was considered a miss. The hit-or-miss score is presented separately in Table 1 for judges and for subjects in the first two data columns. Where information is not supplied for one or the other, the reason is generally that it was impossible for the researchers to obtain it, and for a similar reason the number of cases sometimes varies.[1]

Each data row in Table 1 refers to one segment of the research, and segments for the most part are labeled as they were in the table of Ullman et al. (1973, pp. 275–277). Segments that followed the general procedure I described—all-night sessions, with an agent concentrating on the target during each of the percipient's REM periods—are gathered together in the first eight lines, A through H (in five of these segments, all but A, C, and H, a single percipient continued throughout a series, and in four of these the percipient was a psychologist). Other types of segments are presented in the rest of the table. Lines I, J, and K summarize precognitive sessions; here the target was not selected until after the dreaming and interview had been completed. The target consisted of a set of stimuli to be presented directly to the percipient after it had been selected in the morning. Lines L and M represent GESP sessions in which the percipient's dreams were monitored and recorded throughout the night, but the agent was attempting to transmit only before the percipient went to sleep or just after, or sporadically. Line N refers to a few clairvoyance sessions; these were like the standard GESP sessions except that there was no agent (no one knew the identity of the target). Finally, Line O reports on some GESP sessions in which each dream was considered separately; these formed a single experiment with four percipients, comparing nights involving a different target for each REM period with nights involving repeated use of a single target.

Regardless of the type of session (considering the five types I have described), each session fell into one of two categories: (a) pilot sessions, in which either a new dreamer or a new procedure was being tried out; these appear in lines H, K, and N, or (b) sessions in an experimental series, planned in advance as one or more sessions for each of two or more subjects, or as a number of sessions with the same dreamer throughout. Most of the researchers' publications were devoted to the results obtained in the experimental series, but the results of the pilot sessions have also been briefly reported.

A glance at the score columns for judges and for subjects is sufficient to indicate a strong tendency for an excess of hits over misses. If we average the outcome for judges and for subjects, we find that hits exceed misses on every one of the 15 independent lines on which outcome for hits and misses differs. (On Line E hits and misses occur with equal frequency.) By a simple sign-test, this outcome would be significant beyond the 0.0001 level. I would not stress the exact value here, for several reasons. There was no advance

[1] Of course, usable judgments could not be obtained from the subject in precognitive sessions, because at the time of judging he or she would already know what the target had been. For Line F, the single subject was unable to give the extra time required for judging, and for Line O one of the four subjects failed to make judgments. In a few of the pilot sessions (Lines H, K, and N) only the subject's judgment was sought, and in some sessions only that of one or more judges; in a few the mean judges' rating was neither a hit nor a miss but exactly at the middle.

Table 1
Summary of Maimonides Results on Tendency for Dreams to Be Judged More Like Target Than Like Nontargets in Target Pool

Series	Judges' score		Subjects' score		z or t resulting from judgments		Sources
	Hit	Miss	Hit	Miss	Judges	Subjects	
GESP: Dreams monitored and recorded throughout night; agent "transmitting" during each REM period							
A. 1st screening	7	5	10	2	*z = 0.71*[b]	*z = 1.33*[b]	Ullman, Krippner, & Feldstein (1966)
B. 1st Erwin	5	2	6	1	*z = 2.53*[b]	*z = 1.90*[b]	Ullman et al. (1966)
C. 2nd screening	4	8	9	3	*z = −.25*[b]	*z = 1.17*[b]	Ullman (1969)
D. Posin	6	2	6	2	*z = 1.05*[c]	*z = 1.05*[c]	Ullman (1969)
E. Grayeb	3	5	5	3	*z = −.63*[c]	*z = 0.63*[c]	Ullman, Krippner, & Vaughan (1973)
F. 2nd Erwin	8	0			*t = 4.93*[a]		Ullman & Krippner (1969)
G. Van de Castle	6	2	8	0	*t = 2.81*[a]	*t = 2.74*[a]	Krippner & Ullman (1970)
H. Pilot sessions	53	14	42	22	*z = 4.20*[b]	*z = 2.21*[b]	Ullman et al. (1973)
Precognition: Dreams monitored and recorded throughout night; target experience next day							
I. 1st Bessent	7	1			*t = 2.81*[a]		Krippner, Ullman, & Honorton (1971)
J. 2nd Bessent	7	1			*t = 2.27*[a]		Krippner, Honorton, & Ullman (1972)
K. Pilot sessions	2	0			*z = 0.67*[c]		Ullman et al. (1973)
GESP: Dreams monitored and recorded throughout night; agent active only at beginning or sporadically							
L. Sensory bombardment	8	0	4	4	*z = 3.11*[b]	*z = 0.00*[c]	Krippner, Honorton, Ullman, Masters, & Houston (1971)
M. Grateful Dead	7	5	8	4	*z = 0.61*[c]	*z = 0.81*[c]	Krippner, Honorton, & Ullman (1973)
Clairvoyance: Dreams monitored and recorded throughout night; concealed target known to no one							
N. Pilot sessions	5	3	4	5	*z = 0.98*[b]	*z = 0.00*[b]	Ullman et al. (1973)
GESP: Single dreams							
O. Vaughan, Harris, Parise	105	98	74	79	*z = 0.63*[c]	*z = −.32*[c]	Honorton, Krippner, & Ullman (1972)

Note. GESP = general extrasensory perception. Italics identify results obtained with procedures that preserve independence of judgments in a series. For some series, the published source does not use the uniform measures entered in this table, and mimeographed laboratory reports were also consulted. Superscipts indicate which measure was available, in order of priority.
[a] Ratings. [b] Rankings. [c] Score (count of hits and misses).

plan to merge the outcomes for judges and subjects. Moreover, the various series could be split up in other ways. Although I think my organization of the table is very reasonable (and I did not notice this outcome until after the table was constructed), it is not the organization selected by Ullman et al. (1973); their table, if evaluated statistically in this same way, would not yield so striking a result. What is clear is that the tendency toward hits rather than misses cannot reasonably be ascribed to chance. There is some systematic—that is, nonrandom—source of anomalous resemblance of dreams to target.

Despite its breadth, this "hitting" tendency seems to vary greatly in strength. The data on single dreams—Line O—suggest no consistency. At the other extreme, some separate lines of the table look impressive. I will next consider how we may legitimately evaluate the relative statistical significance of

separate parts of the data on all-night sessions. (I will not try to take exact account here of the fact that the single-dream data are not significant, though it is wise to have in mind that the exact values I cite must be viewed as slightly exaggerated, in the absence of any explicit advance prediction that the results for all-night sessions and for single dreams would differ greatly.)

Two difficulties, one general and one specific, stand in the way of making as thorough an evaluation as I would wish. The general difficulty is that the researchers turned the task of statistical evaluation over to various consultants—for the most part, different consultants at various times—and some of the consultants must also have influenced the choice of procedures and measures. The consultants, and presumably the researchers themselves, seem not to have been at that time very experienced in working with some of the design problems posed by this research nor in planning how the research could be done to permit effective analysis. Much of the research was not properly analyzed at the time, and for much of it the full original data are no longer available. (The researchers have been very helpful in supplying me with material they have been able to locate despite dispersal and storage of the laboratory's files. Perhaps additional details may be recovered in the future.) The result is that completely satisfactory analysis is at present possible only for some portions of the data.

The specific difficulty results from a feature of the research design employed in most of the experimental series, a feature whose implications the researchers did not fully appreciate at the time. If a judge is presented with a set of transcripts and a set of targets and is asked to judge similarity of each target to each transcript, the various judgments may not be completely independent. If one transcript is so closely similar to a particular target that the judge is confident of having recognized a correct match, the judge (or percipient, of course) may minimize the similarity of that target to the transcripts judged later. Instructions to judges explicitly urged them to avoid this error, but we cannot tell how thoroughly this directive was followed. Nonindependence would create no bias toward either positive or negative evidence of correspondence between targets and transcripts, but it would alter variability and thus render inappropriate some standard tests of significance. I have entered in the two succeeding columns of the table a t or a z that can be used in evaluating the statistical significance of the departure from chance expectancy (t is required when ratings are available, and z must be used when only rankings or score counts are available, because sample variability in the former case is estimated from the data but in the latter case must be based conservatively on a theoretical distribution.) If ratings were available, they were used; if not, rankings were used if available; otherwise, score count was used.

Is there likely to have been much of this nonindependence in the series where it was possible? A pertinent fact is that the hits were not generally direct hits. That is, there was no overwhelming tendency for the correct target to be given first place rather than just being ranked in the upper half of the target pool. This greatly reduces the strength of the argument that ordinary significance tests are grossly inaccurate because of nonindependence. Because certainty is not possible, however, we need to separate results according to whether the procedures permitted this kind of nonindependence. In the table, I have italicized results that cannot have been influenced by this difficulty (either because each night's ratings were made by a different person or because each night in a series had, and was judged in relation to, a separate target pool) or that closely approximate this ideal condition.

The outcome is clear. Several segments of the data, considered separately, yield significant evidence that dreams (and associations to them) tended to resemble the picture chosen randomly as target more than they resembled other pictures in the pool. In the case of evaluation by outside judges, two of the three segments that are free of the problem of nonindependence yield separately significant results: The pilot sessions (Line H) yield a z of 4.20, and thus a p of .00002. An experiment with distant but multisensory targets (Line L) yields a z of 3.11 and a p of .001. If we consider segments in which judgments may not be completely independent of each other and analyze them in the standard way, we find that the two series with psychologist William Erwin as dreamer are also significant (if nonindependence of judgments does not seriously interfere), Line B with a z of 2.53 ($p < .01$) and Line F with a t of 4.93 and 7 df ($p < .01$). The two precognitive series (Lines I and J), each with 7 df, yield ts of 2.81 and 2.27, with p values slightly above and below .05, respectively.

Segment results based on the subjects' own judgments of similarity are less significant than those based on judgments by outside judges. Only two segments reach minimal levels of statistical significance: Line G, where the t of 2.74 with 7 df is significant at the .05 level, and Line H, where the z of 2.21 is significant at the .05 level.

The statistical evaluation of the separate segments of the Maimonides experiments also permits a more adequate evaluation of their overall statistical significance. For judgments by outside judges, three segments are free of the potential nonindependence of successive judgments (Lines H, L, and N). Putting these three together by the procedure Mosteller and Bush (1954, pp. 329–330) ascribed to Stouffer (recommended by Rosenthal [1984, p. 72] as the "simplest and most versatile" of the possible procedures), the joint p value is <.000002. For the subjects' own judgments, six segments are available (Lines A, C, G, H,

L, and N), and their joint p value is less than .002. The other segments of the data have the problem of potential nonindependence of successive judgments, and even if the exaggeration of significance may be small for a single line, I would not want to risk compounding it in an overall p. Their prevailing unity of direction, however (direction not being subject to influence by the kind of nonindependence involved here), and the substantial size of some of the differences, justify the inference that the overall evidence of consistency far exceeds that indicated by only those selected segments for which a precise statistical statement is possible. The impression given by the mere count of hits and misses is thus fully confirmed when more sensitive measures are used.

Parapsychological experiments are sometimes criticized on the grounds that what evidence they provide for ESP indicates at most some very small effects detectable only by amassing large bodies of data. Those to whom this criticism has any appeal should be aware that the Maimonides experiments are clearly exempt from it. The significant results on Lines F and G of the table, for example, are each attributable basically to just eight data points.

If replications elsewhere should eventually confirm the statistically significant outcome of the Maimonides experiments, would the fact of statistical significance in itself establish the presence of the kind of anomaly called ESP? Of course not. Statistical significance indicates only the presence of consistency and does not identify its source. ESP, or the more general term *psi,* is a label for consistencies that have no identifiable source and that suggest transfer of information by channels not familiar to present scientific knowledge. A judgment about the appropriateness of the label, and thus about the "ESP hypothesis," is complex. It depends on a variety of other judgments and knowledge—how confidently other possible sources of the consistent effect can be excluded, whether other lines of experimentation are yielding results that suggest the same judgment, and so on.

I believe many psychologists would, like myself, consider the ESP hypothesis to merit serious consideration and continued research if they read the Maimonides reports for themselves and if they familiarized themselves with other recent and older lines of experimentation (e.g., Jahn, 1982, and many of the chapters in Wolman, 1977).

Some parapsychological researchers—among them the Maimonides group—have written at times as though a finding of statistical significance sufficiently justified a conclusion that the apparent anomaly should be classified as ESP. I can understand their choice of words, which is based on their own confidence that their experiments permitted exclusion of other interpretations. But perhaps psychologists who in the future become involved in this area may prefer to use a term such as *anomalies,* so as to avoid variable and possibly confusing connotations about the origin of the anomalies. Zusne and Jones (1982) wisely prepared the way for this usage in speaking of *anomalistic psychology.* But meanwhile, psychologists need not cut themselves off from knowledge of relevant facts because of dissatisfaction with the terminology surrounding their presentation.

Attempted Replications Elsewhere

The Maimonides pattern of controlled experiment in a sleep laboratory, obviously, is extremely time consuming and expensive, and replication seems to have been attempted so far at only two other sleep laboratories. At the University of Wyoming, two experiments yielded results approximately at mean chance expectation—slightly below in one study (Belvedere & Foulkes, 1971), slightly above in the other (Foulkes et al., 1972). In a replication at the Boston University School of Medicine (Globus, Knapp, Skinner, & Healey, 1968), overall results were not significantly positive, though in this instance encouragement for further exploration was reported. The researchers had decided in advance to base their conclusions on exact hits—that is, placing the target first, rather than just in the upper half; by this measure, the results were encouraging, though not statistically significant. Moreover, to quote the researchers, "*Post hoc* analysis revealed that the judges were significantly more correct when they were more 'confident' in their judgments. . . . Further conservatively designed research does seem indicated because of these findings" (Globus et al., 1968, p. 365).

A study by Calvin Hall (1967) is sometimes cited as a replication that confirmed the Maimonides findings; in truth, however, although it provided impressive case material, it was not done in a way that permits evaluation as a replication of the Maimonides experiments. Several small-scale studies, done without the facilities of a sleep laboratory, have been reported that are not replications of even one of the more ambitious Maimonides experiments but each of which reports positive results that might encourage further exploration (Braud, 1977; Child, Kanthamani, & Sweeney, 1977; Rechtschaffen, 1970; Strauch, 1970; Van de Castle, 1971). In the case of these minor studies—unlike the Maimonides studies and the three systematic replications—one must recognize the likelihood of selective publication on the basis of interesting results. Taken all together, these diverse and generally small-scale studies done elsewhere do, in my opinion, add something to the conviction the Maimonides experiments might inspire, that dream research is a promising technique for experimental study of the ESP question.

The lack of significant results in the three systematic replications is hardly conclusive evidence

against eventual replicability. In the Maimonides series, likewise, three successive replications (Lines C, D, and E in Table 1) yielded no significant result, yet they are part of a program yielding highly significant overall results.

If results of such potentially great interest and scientific importance as those of the Maimonides program had been reported on a more conventional topic, one might expect them to be widely and accurately described in reviews of the field to which they were relevant, and to be analyzed carefully as a basis for sound evaluation of whether replication and extension of the research were indicated, or of whether errors could be detected and understood. What has happened in this instance of anomalous research findings?

Representation of the Maimonides Research in Books by Psychologists

It is appropriate to begin with E. M. Hansel's 1980 revision of his earlier critical book on parapsychology. As part of his attempt to bring the earlier book up to date, he included an entire chapter on experiments on telepathy in dreams. One page was devoted to a description of the basic method used in the Maimonides experiments; one paragraph summarized the impressive outcome of 10 of the experiments. The rest of the chapter was devoted mainly to a specific account of the experiment in which psychologist Robert Van de Castle was the subject (the outcome is summarized in Line G of my Table 1) and to the attempted replication at the University of Wyoming (Belvedere & Foulkes, 1971), in which Van de Castle was again the subject. Another page was devoted to another of the Maimonides experiments that was also repeated at the University of Wyoming (Foulkes et al., 1972). Hansel did not mention the replication by Globus et al. (1968), whose authors felt that the results encouraged further exploration. Hansel gave more weight to the two negative outcomes at Wyoming than to the sum of the Maimonides research, arguing that sensory cues supposedly permitted by the procedures at Maimonides, not possible because of greater care taken by the Wyoming experimenters, were responsible for the difference in results. He did not provide, of course, the full account of procedures presented in the original Maimonides reports that might persuade many readers that Hansel's interpretation is far from compelling. Nor did he consider why some of the other experiments at Maimonides, not obviously distinguished in the care with which they were done from the two that were replicated (e.g., those on Lines E, M, and O of Table 1) yielded a close-to-chance outcome such as Hansel might have expected sensory cuing to prevent.

Hansel exaggerated the opportunities for sensory cuing—that is, for the percipient to obtain by ordinary sensory means some information about the target for the night. He did this notably by misinterpreting an ambiguous statement in the Maimonides reports, not mentioning that his interpretation was incompatible with other passages; his interpretation was in fact erroneous, as shown by Akers (1984, pp. 128–129). Furthermore, Hansel did not alert the reader to the great care exerted by the researchers to eliminate possible sources of sensory cuing. Most important is the fact that Hansel did not provide any plausible account—other than fraud—of how the opportunities for sensory cuing that he claimed existed would be likely to lead to the striking findings of the research. For example, he seemed to consider important the fact that at Maimonides the agent could leave his or her room during the night to go to the bathroom, whereas in Wyoming the agent had a room with its own bathroom, and the outer door to the room was sealed with tape to prevent the agent from emerging. Hansel did not attempt to say how the agent's visit to the bathroom could have altered the details of the percipient's dreams each night in a manner distinctively appropriate to that night's target. The only plausible route of influence on the dream record seems to be deliberate fraud involving the researchers and their subjects. The great number and variety of personnel in these studies—experimenters, agents, percipients, and judges—makes fraud especially unlikely as an explanation of the positive findings; but Hansel did not mention this important fact.

It appears to me that all of Hansel's criticisms of the Maimonides experiments are relevant only on the hypothesis of fraud (except for the mistaken criticism I have mentioned above). He said that unintentional communication was more likely but provided no evidence either that it occurred or that such communication—in any form in which it might have occurred—could have produced such consistent results as emerged from the Maimonides experiments. I infer that Hansel was merely avoiding making explicit his unsupported accusations of fraud. Fraud is an interpretation always important to keep in mind, and it is one that could not be entirely excluded even by precautions going beyond those used in the Wyoming studies. But the fact that fraud was as always, theoretically possible hardly justifies dismissal of a series of carefully conducted studies that offer important suggestions for opening up a new line of inquiry into a topic potentially of great significance. Especially regrettable is Hansel's description of various supposed defects in the experiments as though they mark the experiments as being carelessly conducted by general scientific criteria, whereas in fact the supposed defects are relevant only if one assumes fraud. A reader who is introduced to the Maimonides research by Hansel's chapter is likely to get a totally erroneous impression of the care taken by the experimenters to avoid various possible sources of error. The one thing they could

not avoid was obtaining results that Hansel considered a priori impossible, hence evidence of fraud; but Hansel was not entirely frank about his reasoning.

An incidental point worth noting is that Hansel did not himself apply, in his critical attack, the standards of evidence he demanded of the researchers. His conclusions were based implicitly on the assumption that the difference of outcome between the Maimonides and the Wyoming experiments was a genuine difference, not attributable to random variation. He did not even raise the question, as he surely would have if, in some parallel instance, the Maimonides researchers had claimed or implied statistical significance where it was questionable. In fact, the difference of outcome might well have arisen from random error; for the percipient's own judgments the difference is significant at the 5% level (2-tailed), but for the outsiders' judgments it does not approach significance.

Another 1980 book is *The Psychology of Transcendence,* by Andrew Neher, in which almost 100 pages are devoted to "psychic experience." Neher differed from the other authors I refer to in describing the Maimonides work as a "series of studies of great interest" (p. 145), but this evaluation seems to be negated by his devoting only three lines to it and four lines to unsuccessful replications.

A third 1980 publication, *The Psychology of the Psychic,* by David Marks and Richard Kammann, provides less of a general review of recent parapsychology than Hansel's book or even Neher's one long chapter. It is largely devoted to the techniques of mentalists (that is, conjurors specializing in psychological rather than physical effects) and can be useful to anyone encountering a mentalist who pretends to be "psychic." Most readers are not likely to be aware that parapsychological research receives only limited attention. The jacket blurbs give a very different view of the book, as do the authors in their introductory sentences:

> ESP is just around the next corner. When you get there, it is just around the next corner. Having now turned over one hundred of these corners, we decided to call it quits and report our findings for public review. (Marks & Kammann, 1980, p. 4)

Given this introduction to the nature of the book, readers might suppose it would at least mention any corner that many parapsychologists have judged to be an impressive turning. But the Maimonides dream experiments received no mention at all.

Another volume, by psychologist James Alcock (1981), quite clearly purports to include a general review and evaluation of parapsychological research. Alcock mentioned (p. 6) that Hansel had examined the Maimonides experiments, but the only account of them that Alcock offered (on p. 163) was incidental to a discussion of control groups. By implication he seemed to reject the Maimonides experiments because they included no control groups. He wrote that "a control group, for which no sender or no target was used, would appear essential" (p. 163). Later he added, "One could, alternatively, 'send' when the subject was not in the dream state, and compare 'success' in this case with success in dream state trials" (p. 163). The first of these statements suggests a relevant use of control groups but errs in calling it essential; in other psychological research, Alcock would have doubtless readily recognized that within-subject control can, where feasible, be much more efficient and pertinent than a separate control group. His second statement suggests a type of experiment that is probably impossible (because in satisfactory form it seems to require the subject to dream whether awake or asleep and not to know whether he or she was awake or asleep). This second kind of experiment, moreover, has special pertinence only to a comparison between dreaming and waking, not to the question of whether ESP is manifested in dreaming.

Alcock, in short, did not seem to recognize that the design of the Maimonides experiments was based on controls exactly parallel to those used by innumerable psychologists in other research with similar logical structure (and even implied, curiously enough, in his own second suggestion). He encouraged readers to think that the Maimonides studies are beyond the pale of acceptable experimental design, whereas in fact they are fine examples of appropriate use of within-subject control rather than between-subjects control.

The quality of thinking with which Alcock confronted the Maimonides research appeared also in a passage that did not refer to it by name. Referring to an article published in *The Humanist* by Ethel Grodzins Romm, he wrote,

> Romm (1977) argued that a fundamental problem with both the dream telepathy research and the remote viewing tests is that the reports suffer from what she called "shoe-fitting" language; she cited a study in which the sender was installed in a room draped in white fabric and had ice cubes poured down his back. A receiver who reported "white" was immediately judged to have made a "hit" by an independent panel. Yet, as she observed, words such as "miserable", "wet", or "icy" would have been better hits. . . . Again, the obvious need is for a control group. Why are they not used? (p. 163)

What Romm described as "shoe fitting" (misinterpreting events to fit one's expectations) is an important kind of error that is repeatedly made in interpretation of everyday occurrences by people who believe they are psychic. But the dream telepathy research at Maimonides was well protected against this kind of error by the painstaking controls that Alcock seemed not to have noticed. Surely Romm must be referring to some other and very sloppy dream research?

Not at all. The details in this paragraph, and even more in Romm's article, point unmistakably, though inaccurately, to the fifth night of the first precognitive series at Maimonides. The actual details of target and response would alone deprive it of much of its value as an example of shoe fitting. As reported by Krippner, Ullman, & Honorton (1971), the target was a morning experience that included being in a room that was draped with white sheets. The subject's first dream report had included the statement, "I was just standing in a room, surrounded by white. Every imaginable thing in that room was white" (p. 201). There is more similarity here than Romm and Alcock acknowledged in mentioning from this passage only the single word "white."

More important, however, is the fact that the experiment they were referring to provided no opportunity for shoe fitting. The procedures followed in the experiment were completely misrepresented in a way that created the illusion that the possibility existed. There was no panel, in the sense of a group of people gathered together and capable of influencing each other. The judges, operating independently, separately judged every one of the 64 possible combinations of target and transcript yielded by the eight nights of the experiment, not just the eight correct pairings, and they had no clues to which those eight were. Their responses are hardly likely to have been immediate, as they required reading the entire night's transcript. Because each judge was working alone and was not recording times, there would have been no record if a particular response had been immediate, and no record of what particular element in the transcript led to an immediate response.

I looked up in a 1977 issue of *The Humanist* the article by Romm that Alcock cited. The half page on shoe-fitting language gave as examples this item from the Maimonides research and also the SRI remote-viewing experiments (Puthoff & Targ, 1976) done at SRI International. In both cases what was said was pure fiction, based on failure to note what was done in the experiments and in particular that the experimenters were well aware of the danger of shoe-fitting language and that the design of their experiments incorporated procedures to ensure that it could not occur. Romm's ignorance about the Maimonides research and her apparent willingness to fabricate falsehoods about it should be recognized by anyone who had read any of the Maimonides research publications. Yet Alcock accepted and repeated the fictions as though they were true. His presentation in the context of a book apparently in the scientific tradition seems to me more dangerous than Romm's original article, for anyone with a scientific orientation should be able to recognize Romm's article as propaganda. Its title, for example, is "When You Give a Closet Occultist a PhD, What Kind of Research Can You Expect?" and it repeatedly speaks of "cult phuds," meaning people with PhDs who are interested in parapsychological problems. Alcock's repetition of Romm's misstatements in a context lacking these clues may well be taken by many a reader as scholarly writing based on correct information and rational thought. Paradoxically, both Alcock's paragraph and Romm's article are excellent examples of the shoe-fitting error that both decry in others who are in fact carefully avoiding it.

The last of the five books that bring, or fail to bring, the Maimonides research to the attention of psychologists and their students is *Anomalistic Psychology: A Study of Extraordinary Phenomena of Behavior and Experience,* a 1982 volume by Leonard Zusne and Warren H. Jones. This is in many ways an excellent book, and it is also the one of the five that comes closest to including a general review of important recent research in parapsychology. Its brief account of the Maimonides dream experiments, however, misrepresented them in ways that should seriously reduce a reader's interest in considering them further.

Zusne and Jones's description of the basic procedure made three serious errors. First, it implied that one of the experimenters had a chance to know the identity of the target. ("After the subject falls asleep, an art reproduction is selected from a large collection randomly, placed in an envelope, and given to the agent" p. 260). In fact, precautions were taken to ensure that no one but the agent could know the identity of the target. Second, the authors stated that "three judges . . . rate their confidence that the dream content matches the target picture" (p. 260), leading the reader to suppose that the judges were informed of the identity of the target at the time of rating. In fact, a judge was presented with a dream transcript and a pool of potential targets and was asked to rate the degree of similarity between the transcript and each member of the pool, while being unaware of which member had been the target. Third, there was a similarly, though more obscurely, misleading description of how ratings were obtained from the dreamer.

This misinformation was followed by even more serious misrepresentation of the research and, by implication, of the competence of the researchers. Zusne and Jones (1982) wrote that Ullman and Krippner (1978) had found that dreamers were not influenced telepathically unless they knew in advance that an attempt would be made to influence them. This led, they wrote, to the subject's being "primed prior to going to sleep" through the experimenter's

> preparing the receiver through experiences that were related to the content of the picture to be telepathically transmitted during the night. Thus, when the picture was Van Gogh's Corridor of the St. Paul Hospital, which depicts a lonely figure in the hallways of a mental hospital, the receiver: (1)

heard Rosza's *Spellbound* played on a phonograph; (2) heard the monitor laugh hysterically in the room; (3) was addressed as "Mr. Van Gogh" by the monitor; (4) was shown paintings done by mental patients; (5) was given a pill and a glass of water; and (6) was daubed with a piece of cotton dipped in acetone. The receiver was an English "sensitive," but it is obvious that no psychic sensitivity was required to figure out the general content of the picture and to produce an appropriate report, whether any dreams were actually seen or not. (pp. 260–261)

If researchers were to report positive results of the experiment described here by Zusne and Jones and were to claim that it provided some positive evidence of ESP, what would a reader conclude? Surely, that the researchers were completely incompetent, but probably not that they were dishonest. For dishonesty to take such a frank and transparent form is hardly credible.

Incompetence of the researchers is not, however, a proper inference. The simple fact, which anyone can easily verify, is that the account Zusne and Jones gave of the experiment is grossly inaccurate. What Zusne and Jones have done is to describe (for one specific night of the experiment) some of the stimuli provided to the dreamer the next morning, *after* his dreams had been recorded and his night's sleep was over. Zusne and Jones erroneously stated that these stimuli were provided *before* the night's sleep, to prime the subject to have or falsely report having the desired kind of dream. The correct sequence of events was quite clearly stated in the brief reference Zusne and Jones cited (Ullman & Krippner, 1978), as well as in the original research report (Krippner, Honorton, & Ullman, 1972).

I can understand and sympathize with Zusne and Jones's error. The experiment they cited is one in which the nocturnal dreamer was seeking to dream in response to a set of stimuli to be created and presented to him the next morning. As may be seen in Table 1, results from such precognitive sessions (all done with a single subject) were especially strong. This apparent transcendence of time as well as space makes the precognitive findings seem at least doubly impossible to most of us. An easy misreading, therefore, on initially scanning the research report, would be to suppose the stimuli to have been presented partly in advance (because some parts obviously involved a waking subject) and partly during sleep.

This erroneous reading on which Zusne and Jones based their account could easily have been corrected by a more careful rereading. In dealing with other topics, they might have realized the improbability that researchers could have been so grossly incompetent and could have checked the accuracy of their statements before publishing them. Zusne and Jones are not alone in this tendency to quick misperception of parapsychological research through preconception and prejudice; we have already seen it in Alcock's book. Alcock (1983) wrote the review of Zusne and Jones's book for *Contemporary Psychology,* the book-review journal of the American Psychological Association, and he did not mention this egregious error, even though very slight acquaintance with the Maimonides research should suffice to detect it.

Discussion

The experiments at the Maimonides Medical Center on the possibility of ESP in dreams clearly merit careful attention from psychologists who, for whatever reason, are interested in the question of ESP. To firm believers in the impossibility of ESP, they pose a challenge to skill in detecting experimental flaws or to the understanding of other sources of error. To those who can conceive that ESP might be possible, they convey suggestions about some of the conditions influencing its appearance or absence and about techniques for investigating it.

This attention is not likely to be given by psychologists whose knowledge about the experiments comes from the books by their fellow psychologists that purport to review parapsychological research. Some of those books engage in nearly incredible falsification of the facts about the experiments; others simply neglect them. I believe it is fair to say that none of these books has correctly identified any defect in the Maimonides experiments other than ones relevant only to the hypothesis of fraud or on inappropriate statistical reasoning (easily remedied by new calculations from the published data). I do not mean that the Maimonides experiments are models of design and execution. I have already called attention to a design flaw that prevents sensitive analysis of some of the experiments; and the control procedures were violated at one session, as Akers (1984) pointed out on the basis of the full information supplied in the original report. (Neither of these genuine defects was mentioned in any of the five books I have reviewed here, an indication of their authors' general lack of correct information about the Maimonides experiments.)

Readers who doubt that the falsification is as extreme as I have pictured it need only consult the sources I have referred to. Their doubt might also be reduced by familiarity with some of James Bradley's research (1981, 1984). In his 1984 article, he reported similar misrepresentations of fact on a topic, robustness of procedures of statistical inference, on which psychologists would not be thought to have nearly the strength of preconception that many are known to have about ESP. How much more likely, then, falsification on so emotionally laden a topic as ESP is for many psychologists! In the earlier article, Bradley (1981) presented experimental evidence (for college students, in this case, not psychologists) that confi-

dence in the correctness of one's own erroneous opinions is positively correlated with the degree of expertise one believes oneself to have in the field of knowledge within which the erroneous opinion falls. This finding may help in understanding why the authors of some of these books did not find it necessary to consider critically their own erroneous statements.

A very considerable proportion of psychologists have a potential interest in the question of ESP. In a recent survey (Wagner & Monnet, 1979) of university professors in various fields, 34% of psychologists were found to consider ESP either an established fact or a likely possibility, exactly the same proportion as considered it an impossibility. In this survey, psychologists less frequently expressed a positive opinion than did members of other disciplines, a finding that may be attributable to psychologists' better understanding of sources of error in human judgment. There seems to be no equally sound reason for the curious fact that psychologists differed overwhelmingly from others in their tendency to consider ESP an impossibility. Of natural scientists, only 3% checked that opinion; of the 166 professors in other social sciences, not a single one did.

Both of these groups of psychologists have been ill served by the apparently scholarly books that seem to convey information about the dream experiments. The same may be said about some other lines of parapsychological research. Interested readers might well consult the original sources and form their own judgments.

REFERENCES

Akers, C. (1984). Methodological criticisms of parapsychology. In S. Krippner (Ed.), *Advances in parapsychological research* (Vol. 4, pp. 112–164). Jefferson, NC: McFarland.

Alcock, J. E. (1981). *Parapsychology, science or magic? A psychological perspective.* New York: Pergamon Press.

Alcock, J. E. (1983). Bringing anomalies back into psychology. *Contemporary Psychology, 28,* 351–352.

Ayeroff, F., & Abelson, R. P. (1976). ESP and ESB: Belief in personal success at mental telepathy. *Journal of Personality and Social Psychology, 34,* 240–247.

Belvedere, E., & Foulkes, D. (1971). Telepathy and dreams: A failure to replicate. *Perceptual and Motor Skills, 33,* 783–789.

Blackmore, S. J. (1984). A psychological theory of the out-of-body experience. *Journal of Parapsychology, 48,* 201–218.

Bradley, J. V. (1981). Overconfidence in ignorant experts. *Bulletin of the Psychonomic Society, 17,* 82–84.

Bradley, J. V. (1984). Antinonrobustness: A case study in the sociology of science. *Bulletin of the Psychonomic Society, 22,* 463–466.

Braud, W. (1977). Long-distance dream and presleep telepathy. In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.), *Research in parapsychology 1976* (pp. 154–155). Metuchen, NJ: Scarecrow.

Child, I. L., Kanthamani, H., & Sweeney, V. M. (1977). A simplified experiment in dream telepathy. In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.), *Research in parapsychology 1976* (pp. 91–93). Metuchen, NJ: Scarecrow.

Crandall, J. E., & Hite, D. D. (1983). Psi-missing and displacement: Evidence for improperly focused psi? *Journal of the American Society for Psychical Research, 77,* 209–228.

Foulkes, D., Belvedere, E., Masters, R. E. L., Houston, J., Krippner, S., Honorton, C., & Ullman, M. (1972). Long-distance "sensory-bombardment" ESP in dreams: A failure to replicate. *Perceptual and Motor Skills, 35,* 731–734.

Globus, G., Knapp, P., Skinner, J., & Healey, J. (1968). An appraisal of telepathic communication in dreams. *Psychophysiology, 4,* 365.

Hall, C. (1967). Experimente zur telepathischen Beeinflussung von Träumen. [Experiments on telepathic influence on dreams]. *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie, 10,* 18–47.

Hansel, C. E. M. (1980). *ESP and parapsychology: A critical reevaluation.* Buffalo, NY: Prometheus.

Honorton, C., Krippner, S., & Ullman, M. (1972). Telepathic perception of art prints under two conditions. *Proceedings of the 80th Annual Convention of the American Psychological Association, 7,* 319–320.

Hyman, R. (1985). The ganzfeld psi experiment: A critical appraisal. *Journal of Parapsychology, 49,* 3–49.

Irwin, H. J. (1979). *Psi and the mind: An information processing approach.* Metuchen, NJ: Scarecrow.

Jahn, R. G. (1982). The persistent paradox of psychic phenomena: An engineering perspective. *Proceedings of the Institute of Electrical and Electronics Engineers, 70,* 136–170.

Krippner, S., Honorton, E., & Ullman, M. (1972). A second precognitive dream study with Malcolm Bessent. *Journal of the American Society for Psychical Research, 66,* 269–279.

Krippner, S., Honorton, C., & Ullman, M. (1973). An experiment in dream telepathy with "The Grateful Dead." *Journal of the American Society of Psychosomatic Dentistry and Medicine, 20,* 9–17.

Krippner, S., Honorton, C., Ullman, M., Masters, R., & Houston, J. (1971). A long-distance "sensory-bombardment" study of ESP in dreams. *Journal of the American Society for Psychical Research, 65,* 468–475.

Krippner, S., & Ullman, M. (1970). Telepathy and dreams: A controlled experiment with electroencephalogram-electro-oculogram monitoring. *Journal of Nervous and Mental Disease, 151,* 394–403.

Krippner, S., Ullman, M., & Honorton, C. (1971). A precognitive dream study with a single subject. *Journal of the American Society for Psychical Research, 65,* 192–203.

Lowry, R. (1981). Apparent PK effect on computer-generated random digit series. *Journal of the American Society for Psychical Research, 75,* 209–220.

Marks, D., & Kammann, R. (1980). *The psychology of the psychic.* Buffalo, NY: Prometheus.

Mosteller, F., & Bush, R. R. (1954). Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 1, pp. 289–334). Cambridge, MA: Addison-Wesley.

Neher, A. (1980). *The psychology of transcendence.* Englewood Cliffs, NJ: Prentice-Hall.

Puthoff, H. E., & Targ, R. (1976). A perceptual channel for information transfer over kilometer distances: Historical perspective and recent research. *Proceedings of the Institute of Electrical and Electronic Engineers, 64,* 329–354.

Radin, D. I. (1982). Experimental attempts to influence pseudo-random number sequences. *Journal of the American Society for Psychical Research, 76,* 359–374.

Rechtschaffen, A. (1970). Sleep and dream states: An experimental design. In R. Cavanna (Ed.), *Psi favorable states of consciousness* (pp. 87–120). New York: Parapsychology Foundation.

Romm, E. G. (1977). When you give a closet occultist a Ph.D., what kind of research can you expect? *The Humanist, 37*(3), 12–15.

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage.

Strauch, I. (1970). Dreams and psi in the laboratory. In R. Cavanna

(Ed.), *Psi favorable states of consciousness* (pp. 46–54). New York: Parapsychology Foundation.

Ullman, M. (1969). Telepathy and dreams. *Experimental Medicine & Surgery, 27,* 19–38.

Ullman, M., & Krippner, S. (1969). A laboratory approach to the nocturnal dimension of paranormal experience: Report of a confirmatory study using the REM monitoring technique. *Biological Psychiatry, 1,* 259–270.

Ullman, M., & Krippner, S. (1978). Experimental dream studies. In M. Ebon (Ed.), *The Signet handbook of parapsychology* (pp. 409–422). New York: New American Library.

Ullman, M., Krippner, S., & Feldstein, S. (1966). Experimentally induced telepathic dreams: Two studies using EEG-REM monitoring technique. *International Journal of Neuropsychiatry, 2,* 420–437.

Ullman, M., Krippner, S., & Vaughan, A. (1973). *Dream telepathy.* New York: Macmillan.

Van de Castle, R. L. (1971). The study of GESP in a group setting by means of dreams. *Journal of Parapsychology, 35,* 312.

Wagner, M. W., & Monnet, M. (1979). Attitudes of college professors toward extra-sensory perception. *Zetetic Scholar,* no. 5, 7–16.

Wolman, B. B. (Ed.). (1977). *Handbook of parapsychology.* New York: Van Nostrand Reinhold.

Zusne, L., & Jones, W. H. (1982). *Anomalistic psychology: A study of extraordinary phenomena of behavior and experience.* Hillsdale, NJ: Erlbaum.

[16]

Does Psi Exist? Replicable Evidence for an Anomalous Process of Information Transfer

Daryl J. Bem and Charles Honorton

Most academic psychologists do not yet accept the existence of psi, anomalous processes of information or energy transfer (such as telepathy or other forms of extrasensory perception) that are currently unexplained in terms of known physical or biological mechanisms. We believe that the replication rates and effect sizes achieved by one particular experimental method, the ganzfeld procedure, are now sufficient to warrant bringing this body of data to the attention of the wider psychological community. Competing meta-analyses of the ganzfeld database are reviewed, 1 by R. Hyman (1985), a skeptical critic of psi research, and the other by C. Honorton (1985), a parapsychologist and major contributor to the ganzfeld database. Next the results of 11 new ganzfeld studies that comply with guidelines jointly authored by R. Hyman and C. Honorton (1986) are summarized. Finally, issues of replication and theoretical explanation are discussed.

The term *psi* denotes anomalous processes of information or energy transfer, processes such as telepathy or other forms of extrasensory perception that are currently unexplained in terms of known physical or biological mechanisms. The term is purely descriptive: It neither implies that such anomalous phenomena are paranormal nor connotes anything about their underlying mechanisms.

Does psi exist? Most academic psychologists don't think so. A survey of more than 1,100 college professors in the United States found that 55% of natural scientists, 66% of social scientists (excluding psychologists), and 77% of academics in the arts, humanities, and education believed that ESP is either an established fact or a likely possibility. The comparable figure for psychologists was only 34%. Moreover, an equal number of psychologists declared ESP to be an impossibility, a view expressed by only 2% of all other respondents (Wagner & Monnet, 1979).

We psychologists are probably more skeptical about psi for several reasons. First, we believe that extraordinary claims require extraordinary proof. And although our colleagues from other disciplines would probably agree with this dictum, we are more likely to be familiar with the methodological and statistical requirements for sustaining such claims, as well as with previous claims that failed either to meet those requirements or to survive the test of successful replication. Even for ordinary claims, our conventional statistical criteria are conservative. The sacred $p = .05$ threshold is a constant reminder that it is far more sinful to assert that an effect exists when it does not (the Type I error) than to assert that an effect does not exist when it does (the Type II error).

Second, most of us distinguish sharply between phenomena whose explanations are merely obscure or controversial (e.g., hypnosis) and phenomena such as psi that appear to fall outside our current explanatory framework altogether. (Some would characterize this as the difference between the unexplained and the inexplicable.) In contrast, many laypersons treat all exotic psychological phenomena as epistemologically equivalent; many even consider déjà vu to be a psychic phenomenon. The blurring of this critical distinction is aided and abetted by the mass media, "new age" books and mind-power courses, and "psychic" entertainers who present both genuine hypnosis and fake "mind reading" in the course of a single performance. Accordingly, most laypersons would not have to revise their conceptual model of reality as radically as we would in order to assimilate the existence of psi. For us, psi is simply more extraordinary.

Finally, research in cognitive and social psychology has sensitized us to the errors and biases that plague intuitive attempts to draw valid inferences from the data of everyday experience (Gilovich, 1991; Nisbett & Ross, 1980; Tversky & Kahneman, 1971). This leads us to give virtually no probative weight to anecdotal or journalistic reports of psi, the main source cited by

---

Daryl J. Bem, Department of Psychology, Cornell University; Charles Honorton, Department of Psychology, University of Edinburgh, Edinburgh, Scotland.

Sadly, Charles Honorton died of a heart attack on November 4, 1992, 9 days before this article was accepted for publication. He was 46. Parapsychology has lost one of its most valued contributors. I have lost a valued friend.

This collaboration had its origins in a 1983 visit I made to Honorton's Psychophysical Research Laboratories (PRL) in Princeton, New Jersey, as one of several outside consultants brought in to examine the design and implementation of the experimental protocols.

Preparation of this article was supported, in part, by grants to Charles Honorton from the American Society for Psychical Research and the Parapsychology Foundation, both of New York City. The work at PRL summarized in the second half of this article was supported by the James S. McDonnell Foundation of St. Louis, Missouri, and by the John E. Fetzer Foundation of Kalamazoo, Michigan.

Helpful comments on drafts of this article were received from Deborah Delanoy, Edwin May, Donald McCarthy, Robert Morris, John Palmer, Robert Rosenthal, Lee Ross, Jessica Utts, Philip Zimbardo, and two anonymous reviewers.

Correspondence concerning this article should be addressed to Daryl J. Bem, Department of Psychology, Uris Hall, Cornell University, Ithaca, New York 14853. Electronic mail may be sent to d.bem@cornell.edu.

our academic colleagues as evidence for their beliefs about psi (Wagner & Monnet, 1979).

Ironically, however, psychologists are probably not more familiar than others with recent experimental research on psi. Like most psychological research, parapsychological research is reported primarily in specialized journals; unlike most psychological research, however, contemporary parapsychological research is not usually reviewed or summarized in psychology's textbooks, handbooks, or mainstream journals. For example, only 1 of 64 introductory psychology textbooks recently surveyed even mentions the experimental procedure reviewed in this article, a procedure that has been in widespread use since the early 1970s (Roig, Icochea, & Cuzzucoli, 1991). Other secondary sources for nonspecialists are frequently inaccurate in their descriptions of parapsychological research. (For discussions of this problem, see Child, 1985, and Palmer, Honorton, & Utts, 1989.)

This situation may be changing. Discussions of modern psi research have recently appeared in a widely used introductory textbook (Atkinson, Atkinson, Smith, & Bem, 1990, 1993), two mainstream psychology journals (Child, 1985; Rao & Palmer, 1987), and a scholarly but accessible book for nonspecialists (Broughton, 1991). The purpose of the present article is to supplement these broader treatments with a more detailed, meta-analytic presentation of evidence issuing from a single experimental method: the *ganzfeld* procedure. We believe that the replication rates and effect sizes achieved with this procedure are now sufficient to warrant bringing this body of data to the attention of the wider psychological community.

The Ganzfeld Procedure

By the 1960s, a number of parapsychologists had become dissatisfied with the familiar ESP testing methods pioneered by J. B. Rhine at Duke University in the 1930s. In particular, they believed that the repetitive forced-choice procedure in which a subject repeatedly attempts to select the correct "target" symbol from a set of fixed alternatives failed to capture the circumstances that characterize reported instances of psi in everyday life.

Historically, psi has often been associated with meditation, hypnosis, dreaming, and other naturally occurring or deliberately induced altered states of consciousness. For example, the view that psi phenomena can occur during meditation is expressed in most classical texts on meditative techniques; the belief that hypnosis is a psi-conducive state dates all the way back to the days of early mesmerism (Dingwall, 1968); and cross-cultural surveys indicate that most reported "real-life" psi experiences are mediated through dreams (Green, 1960; Prasad & Stevenson, 1968; L. E. Rhine, 1962; Sannwald, 1959).

There are now reports of experimental evidence consistent with these anecdotal observations. For example, several laboratory investigators have reported that meditation facilitates psi performance (Honorton, 1977). A meta-analysis of 25 experiments on hypnosis and psi conducted between 1945 and 1981 in 10 different laboratories suggests that hypnotic induction may also facilitate psi performance (Schechter, 1984). And dream-mediated psi was reported in a series of experiments conducted at Maimonides Medical Center in New York and published between 1966 and 1972 (Child, 1985; Ullman, Krippner, & Vaughan, 1973).

In the Maimonides dream studies, two subjects—a "receiver" and a "sender"—spent the night in a sleep laboratory. The receiver's brain waves and eye movements were monitored as he or she slept in an isolated room. When the receiver entered a period of REM sleep, the experimenter pressed a buzzer that signaled the sender—under the supervision of a second experimenter—to begin a sending period. The sender would then concentrate on a randomly chosen picture (the "target") with the goal of influencing the content of the receiver's dream.

Toward the end of the REM period, the receiver was awakened and asked to describe any dream just experienced. This procedure was repeated throughout the night with the same target. A transcription of the receiver's dream reports was given to outside judges who blindly rated the similarity of the night's dreams to several pictures, including the target. In some studies, similarity ratings were also obtained from the receivers themselves. Across several variations of the procedure, dreams were judged to be significantly more similar to the target pictures than to the control pictures in the judging sets (failures to replicate the Maimonides results were also reviewed by Child, 1985).

These several lines of evidence suggested a working model of psi in which psi-mediated information is conceptualized as a weak signal that is normally masked by internal somatic and external sensory "noise." By reducing ordinary sensory input, these diverse psi-conducive states are presumed to raise the signal-to-noise ratio, thereby enhancing a person's ability to detect the psi-mediated information (Honorton, 1969, 1977). To test the hypothesis that a reduction of sensory input itself facilitates psi performance, investigators turned to the ganzfeld procedure (Braud, Wood, & Braud, 1975; Honorton & Harper, 1974; Parker, 1975), a procedure originally introduced into experimental psychology during the 1930s to test propositions derived from gestalt theory (Avant, 1965; Metzger, 1930).

Like the dream studies, the psi ganzfeld procedure has most often been used to test for telepathic communication between a sender and a receiver. The receiver is placed in a reclining chair in an acoustically isolated room. Translucent ping-pong ball halves are taped over the eyes and headphones are placed over the ears; a red floodlight directed toward the eyes produces an undifferentiated visual field, and white noise played through the headphones produces an analogous auditory field. It is this homogeneous perceptual environment that is called the *Ganzfeld* ("total field"). To reduce internal somatic "noise," the receiver typically also undergoes a series of progressive relaxation exercises at the beginning of the ganzfeld period.

The sender is sequestered in a separate acoustically isolated room, and a visual stimulus (art print, photograph, or brief videotaped sequence) is randomly selected from a large pool of such stimuli to serve as the target for the session. While the sender concentrates on the target, the receiver provides a continuous verbal report of his or her ongoing imagery and mentation, usually for about 30 minutes. At the completion of the ganzfeld period, the receiver is presented with several stimuli (usually four) and, without knowing which stimulus was the target, is asked to rate the degree to which each matches the imagery and mentation experienced during the ganzfeld period. If the receiver assigns the highest rating to the target stimulus, it

is scored as a "hit." Thus, if the experiment uses judging sets containing four stimuli (the target and three decoys or control stimuli), the hit rate expected by chance is .25. The ratings can also be analyzed in other ways; for example, they can be converted to ranks or standardized scores within each set and analyzed parametrically across sessions. And, as with the dream studies, the similarity ratings can also be made by outside judges using transcripts of the receiver's mentation report.

Meta-Analyses of the Ganzfeld Database

In 1985 and 1986, the *Journal of Parapsychology* devoted two entire issues to a critical examination of the ganzfeld database. The 1985 issue comprised two contributions: (a) a meta-analysis and critique by Ray Hyman (1985), a cognitive psychologist and skeptical critic of parapsychological research, and (b) a competing meta-analysis and rejoinder by Charles Honorton (1985), a parapsychologist and major contributor to the ganzfeld database. The 1986 issue contained four commentaries on the Hyman–Honorton exchange, a joint communiqué by Hyman and Honorton, and six additional commentaries on the joint communiqué itself. We summarize the major issues and conclusions here.

Replication Rates

Rates by study. Hyman's meta-analysis covered 42 psi ganzfeld studies reported in 34 separate reports written or published from 1974 through 1981. One of the first problems he discovered in the database was multiple analysis. As noted earlier, it is possible to calculate several indexes of psi performance in a ganzfeld experiment and, furthermore, to subject those indexes to several kinds of statistical treatment. Many investigators reported multiple indexes or applied multiple statistical tests without adjusting the criterion significance level for the number of tests conducted. Worse, some may have "shopped" among the alternatives until finding one that yielded a significantly successful outcome. Honorton agreed that this was a problem.

Accordingly, Honorton applied a uniform test on a common index across all studies from which the pertinent datum could be extracted, regardless of how the investigators had analyzed the data in the original reports. He selected the proportion of hits as the common index because it could be calculated for the largest subset of studies: 28 of the 42 studies. The hit rate is also a conservative index because it discards most of the rating information; a second place ranking—a "near miss"—receives no more credit than a last place ranking. Honorton then calculated the exact binomial probability and its associated z score for each study.

Of the 28 studies, 23 (82%) had positive z scores ($p = 4.6 \times 10^{-4}$, exact binomial test with $p = q = .5$). Twelve of the studies (43%) had z scores that were independently significant at the 5% level ($p = 3.5 \times 10^{-9}$, binomial test with 28 studies, $p = .05$, and $q = .95$), and 7 of the studies (25%) were independently significant at the 1% level ($p = 9.8 \times 10^{-9}$). The composite Stouffer z score across the 28 studies was 6.60 ($p = 2.1 \times 10^{-11}$).[1] A more conservative estimate of significance can be obtained by including 10 additional studies that also used the relevant judging procedure but did not report hit rates. If these studies are assigned a mean z score of zero, the Stouffer z across all 38 studies becomes 5.67 ($p = 7.3 \times 10^{-9}$).

Thus, whether one considers only the studies for which the relevant information is available or includes a null estimate for the additional studies for which the information is not available, the aggregate results cannot reasonably be attributed to chance. And, by design, the cumulative outcome reported here cannot be attributed to the inflation of significance levels through multiple analysis.

Rates by laboratory. One objection to estimates such as those just described is that studies from a common laboratory are not independent of one another (Parker, 1978). Thus, it is possible for one or two investigators to be disproportionately responsible for a high replication rate, whereas other, independent investigators are unable to obtain the effect.

The ganzfeld database is vulnerable to this possibility. The 28 studies providing hit rate information were conducted by investigators in 10 different laboratories. One laboratory contributed 9 of the studies, Honorton's own laboratory contributed 5, 2 other laboratories contributed 3 each, 2 contributed 2 each, and the remaining 4 laboratories each contributed 1. Thus, half of the studies were conducted by only 2 laboratories, 1 of them Honorton's own.

Accordingly, Honorton calculated a separate Stouffer z score for each laboratory. Significantly positive outcomes were reported by 6 of the 10 laboratories, and the combined z score across laboratories was 6.16 ($p = 3.6 \times 10^{-10}$). Even if all of the studies conducted by the 2 most prolific laboratories are discarded from the analysis, the Stouffer z across the 8 other laboratories remains significant ($z = 3.67$, $p = 1.2 \times 10^{-4}$). Four of these studies are significant at the 1% level ($p = 9.2 \times 10^{-6}$, binomial test with 14 studies, $p = .01$, and $q = .99$), and each was contributed by a different laboratory. Thus, even though the total number of laboratories in this database is small, most of them have reported significant studies, and the significance of the overall effect does not depend on just one or two of them.

Selective Reporting

In recent years, behavioral scientists have become increasingly aware of the "file-drawer" problem: the likelihood that successful studies are more likely to be published than unsuccessful studies, which are more likely to be consigned to the file drawers of their disappointed investigators (Bozarth & Roberts, 1972; Sterling, 1959). Parapsychologists were among the first to become sensitive to the problem, and, in 1975, the Parapsychological Association Council adopted a policy opposing the selective reporting of positive outcomes. As a consequence, negative findings have been routinely reported at the association's meetings and in its affiliated publications for almost two decades. As has already been shown, more than half of the ganzfeld studies included in the meta-analysis yielded outcomes whose significance falls short of the conventional .05 level.

A variant of the selective reporting problem arises from what

[1] Stouffer's z is computed by dividing the sum of the z scores for the individual studies by the square root of the number of studies (Rosenthal, 1978).

Hyman (1985) has termed the "retrospective study." An investigator conducts a small set of exploratory trials. If they yield null results, they remain exploratory and never become part of the official record; if they yield positive results, they are defined as a study after the fact and are submitted for publication. In support of this possibility, Hyman noted that there are more significant studies in the database with fewer than 20 trials than one would expect under the assumption that, all other things being equal, statistical power should increase with the square root of the sample size. Although Honorton questioned the assumption that "all other things" are in fact equal across the studies and disagreed with Hyman's particular statistical analysis, he agreed that there is an apparent clustering of significant studies with fewer than 20 trials. (Of the complete ganzfeld database of 42 studies, 8 involved fewer than 20 trials, and 6 of those studies reported statistically significant results.)

Because it is impossible, by definition, to know how many unknown studies—exploratory or otherwise—are languishing in file drawers, the major tool for estimating the seriousness of selective reporting problems has become some variant of Rosenthal's file-drawer statistic, an estimate of how many unreported studies with z scores of zero would be required to exactly cancel out the significance of the known database (Rosenthal, 1979). For the 28 direct-hit ganzfeld studies alone, this estimate is 423 fugitive studies, a ratio of unreported-to-reported studies of approximately 15:1. When it is recalled that a single ganzfeld session takes over an hour to conduct, it is not surprising that—despite his concern with the retrospective study problem—Hyman concurred with Honorton and other participants in the published debate that selective reporting cannot plausibly account for the overall statistical significance of the psi ganzfeld database (Hyman & Honorton, 1986).[2]

Methodological Flaws

If the most frequent criticism of parapsychology is that it has not produced a replicable psi effect, the second most frequent criticism is that many, if not most, psi experiments have inadequate controls and procedural safeguards. A frequent charge is that positive results emerge primarily from initial, poorly controlled studies and then vanish as better controls and safeguards are introduced.

Fortunately, meta-analysis provides a vehicle for empirically evaluating the extent to which methodological flaws may have contributed to artifactual positive outcomes across a set of studies. First, ratings are assigned to each study that index the degree to which particular methodological flaws are or are not present; these ratings are then correlated with the studies' outcomes. Large positive correlations constitute evidence that the observed effect may be artifactual.

In psi research, the most fatal flaws are those that might permit a subject to obtain the target information in normal sensory fashion, either inadvertently or through deliberate cheating. This is called the problem of *sensory leakage*. Another potentially serious flaw is inadequate randomization of target selection.

Sensory leakage. Because the ganzfeld is itself a perceptual isolation procedure, it goes a long way toward eliminating potential sensory leakage during the ganzfeld portion of the session. There are, however, potential channels of sensory leakage after the ganzfeld period. For example, if the experimenter who interacts with the receiver knows the identity of the target, he or she could bias the receiver's similarity ratings in favor of correct identification. Only one study in the database contained this flaw, a study in which subjects actually performed slightly below chance expectation. Second, if the stimulus set given to the receiver for judging contains the actual physical target handled by the sender during the sending period, there might be cues (e.g., fingerprints, smudges, or temperature differences) that could differentiate the target from the decoys. Moreover, the process of transferring the stimulus materials to the receiver's room itself opens up other potential channels of sensory leakage. Although contemporary ganzfeld studies have eliminated both of these possibilities by using duplicate stimulus sets, some of the earlier studies did not.

Independent analyses by Hyman and Honorton agreed that there was no correlation between inadequacies of security against sensory leakage and study outcome. Honorton further reported that if studies that failed to use duplicate stimulus sets were discarded from the analysis, the remaining studies are still highly significant (Stouffer $z = 4.35$, $p = 6.8 \times 10^{-6}$).

Randomization. In many psi experiments, the issue of target randomization is critical because systematic patterns in inadequately randomized target sequences might be detected by subjects during a session or might match subjects' preexisting response biases. In a ganzfeld study, however, randomization is a much less critical issue because only one target is selected during the session and most subjects serve in only one session. The primary concern is simply that all the stimuli within each judging set be sampled uniformly over the course of the study. Similar considerations govern the second randomization, which takes place after the ganzfeld period and determines the sequence in which the target and decoys are presented to the receiver (or external judge) for judging.

Nevertheless, Hyman and Honorton disagreed over the findings here. Hyman claimed there was a correlation between flaws of randomization and study outcome; Honorton claimed there was not. The sources of this disagreement were in conflicting definitions of flaw categories, in the coding and assignment of flaw ratings to individual studies, and in the subsequent statistical treatment of those ratings.

Unfortunately, there have been no ratings of flaws by independent raters who were unaware of the studies' outcomes (Morris, 1991). Nevertheless, none of the contributors to the subsequent debate concurred with Hyman's conclusion, whereas four nonparapsychologists—two statisticians and two psychologists—explicitly concurred with Honorton's conclusion (Harris & Rosenthal, 1988b; Saunders, 1985; Utts, 1991a). For example, Harris and Rosenthal (one of the pioneers in the use of meta-analysis in psychology) used Hyman's own flaw ratings and failed to find any significant relationships between flaws and study outcomes in each of two separate analyses:

[2] A 1980 survey of parapsychologists uncovered only 19 completed but unreported ganzfeld studies. Seven of these had achieved significantly positive results, a proportion (.37) very similar to the proportion of independently significant studies in the meta-analysis (.43) (Blackmore, 1980).

"Our analysis of the effects of flaws on study outcome lends no support to the hypothesis that Ganzfeld research results are a significant function of the set of flaw variables" (1988b, p. 3; for a more recent exchange regarding Hyman's analysis, see Hyman, 1991; Utts, 1991a, 1991b).

Effect Size

Some critics of parapsychology have argued that even if current laboratory-produced psi effects turn out to be replicable and nonartifactual, they are too small to be of theoretical interest or practical importance. We do not believe this to be the case for the psi ganzfeld effect.

In psi ganzfeld studies, the hit rate itself provides a straightforward descriptive measure of effect size, but this measure cannot be compared directly across studies because they do not all use a four-stimulus judging set and, hence, do not all have a chance baseline of .25. The next most obvious candidate, the difference in each study between the hit rate observed and the hit rate expected under the null hypothesis, is also intuitively descriptive but is not appropriate for statistical analysis because not all differences between proportions that are equal are equally detectable (e.g., the power to detect the difference between .55 and .25 is different from the power to detect the difference between .50 and .20).

To provide a scale of equal detectability, Cohen (1988) devised the effect size index h, which involves an arcsine transformation on the proportions before calculation of their difference. Cohen's h is quite general and can assess the difference between any two proportions drawn from independent samples or between a single proportion and any specified hypothetical value. For the 28 studies examined in the meta-analyses, h was .28, with a 95% confidence interval from .11 to .45.

But because values of h do not provide an intuitively descriptive scale, Rosenthal and Rubin (1989; Rosenthal, 1991) have recently suggested a new index, π, which applies specifically to one-sample, multiple-choice data of the kind obtained in ganzfeld experiments. In particular, π expresses all hit rates as the proportion of hits that would have been obtained if there had been only two equally likely alternatives—essentially a coin flip. Thus, π ranges from 0 to 1, with .5 expected under the null hypothesis. The formula is

$$\pi = \frac{P(k-1)}{P(k-2)+1},$$

where P is the raw proportion of hits and k is the number of alternative choices available. Because π has such a straightforward intuitive interpretation, we use it (or its conversion back to an equivalent four-alternative hit rate) throughout this article whenever it is applicable.

For the 28 studies examined in the meta-analyses, the mean value of π was .62, with a 95% confidence interval from .55 to .69. This corresponds to a four-alternative hit rate of 35%, with a 95% confidence interval from 28% to 43%.

Cohen (1988, 1992) has also categorized effect sizes into small, medium, and large, with medium denoting an effect size that should be apparent to the naked eye of a careful observer. For a statistic such as π, which indexes the deviation of a proportion from .5, Cohen considers .65 to be a medium effect size: A statistically unaided observer should be able to detect the bias of a coin that comes up heads on 65% of the trials. Thus, at .62, the psi ganzfeld effect size falls just short of Cohen's naked-eye criterion. From the phenomenology of the ganzfeld experimenter, the corresponding hit rate of 35% implies that he or she will see a subject obtain a hit approximately every third session rather than every fourth.

It is also instructive to compare the psi ganzfeld effect with the results of a recent medical study that sought to determine whether aspirin can prevent heart attacks (Steering Committee of the Physicians' Health Study Research Group, 1988). The study was discontinued after 6 years because it was already clear that the aspirin treatment was effective ($p < .00001$) and it was considered unethical to keep the control group on placebo medication. The study was widely publicized as a major medical breakthrough. But despite its undisputed reality and practical importance, the size of the aspirin effect is quite small: Taking aspirin reduces the probability of suffering a heart attack by only .008. The corresponding effect size (h) is .068, about one third to one fourth the size of the psi ganzfeld effect (Atkinson et al., 1993, p. 236; Utts, 1991b).

In sum, we believe that the psi ganzfeld effect is large enough to be of both theoretical interest and potential practical importance.

Experimental Correlates of the Psi Ganzfeld Effect

We showed earlier that the technique of correlating variables with effect sizes across studies can help to assess whether methodological flaws might have produced artifactual positive outcomes. The same technique can be used more affirmatively to explore whether an effect varies systematically with conceptually relevant variations in experimental procedure. The discovery of such correlates can help to establish an effect as genuine, suggest ways of increasing replication rates and effect sizes, and enhance the chances of moving beyond the simple demonstration of an effect to its explanation. This strategy is only heuristic, however. Any correlates discovered must be considered quite tentative, both because they emerge from post hoc exploration and because they necessarily involve comparisons across heterogeneous studies that differ simultaneously on many interrelated variables, known and unknown. Two such correlates emerged from the meta-analyses of the psi ganzfeld effect.

Single-versus multiple-image targets. Although most of the 28 studies in the meta-analysis used single pictures as targets, 9 (conducted by three different investigators) used View Master stereoscopic slide reels that presented multiple images focused on a central theme. Studies using the View Master reels produced significantly higher hit rates than did studies using the single-image targets (50% vs. 34%), $t(26) = 2.22$, $p = .035$, two-tailed.

Sender–receiver pairing. In 17 of the 28 studies, participants were free to bring in friends to serve as senders. In 8 studies, only laboratory-assigned senders were used. (Three studies used no sender.) Unfortunately, there is no record of how many participants in the former studies actually brought in friends. Nevertheless, those 17 studies (conducted by six different investigators) had significantly higher hit rates than did the studies

that used only laboratory-assigned senders (44% vs. 26%), $t(23) = 2.39$, $p = .025$, two-tailed.

The Joint Communiqué

After their published exchange in 1985, Hyman and Honorton agreed to contribute a joint communiqué to the subsequent discussion that was published in 1986. First, they set forth their areas of agreement and disagreement:

> We agree that there is an overall significant effect in this data base that cannot reasonably be explained by selective reporting or multiple analysis. We continue to differ over the degree to which the effect constitutes evidence for psi, but we agree that the final verdict awaits the outcome of future experiments conducted by a broader range of investigators and according to more stringent standards. (Hyman & Honorton, 1986, p. 351)

They then spelled out in detail the "more stringent standards" they believed should govern future experiments. These standards included strict security precautions against sensory leakage, testing and documentation of randomization methods for selecting targets and sequencing the judging pool, statistical correction for multiple analyses, advance specification of the status of the experiment (e.g., pilot study or confirmatory experiment), and full documentation in the published report of the experimental procedures and the status of statistical tests (e.g., planned or post hoc).

The National Research Council Report

In 1988, the National Research Council (NRC) of the National Academy of Sciences released a widely publicized report commissioned by the U.S. Army that assessed several controversial technologies for enhancing human performance, including accelerated learning, neurolinguistic programming, mental practice, biofeedback, and parapsychology (Druckman & Swets, 1988; summarized in Swets & Bjork, 1990). The report's conclusion concerning parapsychology was quite negative: "The Committee finds no scientific justification from research conducted over a period of 130 years for the existence of parapsychological phenomena" (Druckman & Swets, 1988, p. 22).

An extended refutation strongly protesting the committee's treatment of parapsychology has been published elsewhere (Palmer et al., 1989). The pertinent point here is simply that the NRC's evaluation of the ganzfeld studies does not reflect an additional, independent examination of the ganzfeld database but is based on the same meta-analysis conducted by Hyman that we have discussed in this article.

Hyman chaired the NRC's Subcommittee on Parapsychology, and, although he had concurred with Honorton 2 years earlier in their joint communiqué that "there is an overall significant effect in this data base that cannot reasonably be explained by selective reporting or multiple analysis" (p. 351) and that "significant outcomes have been produced by a number of different investigators" (p. 352), neither of these points is acknowledged in the committee's report.

The NRC also solicited a background report from Harris and Rosenthal (1988a), which provided the committee with a comparative methodological analysis of the five controversial areas just listed. Harris and Rosenthal noted that, of these areas, "only the Ganzfeld ESP studies [the only psi studies they evaluated] regularly meet the basic requirements of sound experimental design" (p. 53), and they concluded that

> it would be implausible to entertain the null given the combined p from these 28 studies. Given the various problems or flaws pointed out by Hyman and Honorton . . . we might estimate the obtained accuracy rate to be about 1/3 . . . when the accuracy rate expected under the null is 1/4. (p. 51)[3]

The Autoganzfeld Studies

In 1983, Honorton and his colleagues initiated a new series of ganzfeld studies designed to avoid the methodological problems he and others had identified in earlier studies (Honorton, 1979; Kennedy, 1979). These studies complied with all of the detailed guidelines that he and Hyman were to publish later in their joint communiqué. The program continued until September 1989, when a loss of funding forced the laboratory to close. The major innovations of the new studies were computer control of the experimental protocol—hence the name *autoganzfeld*—and the introduction of videotaped film clips as target stimuli.

Method

The basic design of the autoganzfeld studies was the same as that described earlier[4]: A receiver and sender were sequestered in separate, acoustically isolated chambers. After a 14-min period of progressive relaxation, the receiver underwent ganzfeld stimulation while describing his or her thoughts and images aloud for 30 min. Meanwhile, the sender concentrated on a randomly selected target. At the end of the ganzfeld period, the receiver was shown four stimuli and, without knowing which of the four had been the target, rated each stimulus for its similarity to his or her mentation during the ganzfeld.

The targets consisted of 80 still pictures (static targets) and 80 short video segments complete with soundtracks (dynamic targets), all recorded on videocassette. The static targets included art prints, photographs, and magazine advertisements; the dynamic targets included excerpts of approximately 1-min duration from motion pictures, TV shows, and cartoons. The 160 targets were arranged in judging sets of four static or four dynamic targets each, constructed to minimize similarities among targets within a set.

Target selection and presentation. The VCR containing the taped targets was interfaced to the controlling computer, which selected the target and controlled its repeated presentation to the sender during the ganzfeld period, thus eliminating the need for a second experimenter to accompany the sender. After the ganzfeld period, the computer randomly sequenced the four-clip judging set and presented it to the receiver on a TV monitor for judging. The receiver used a computer game paddle to make his or her ratings on a 40-point scale that appeared on

---

[3] In a troubling development, the chair of the NRC Committee phoned Rosenthal and asked him to delete the parapsychology section of the paper (R. Rosenthal, personal communication, September 15, 1992). Although Rosenthal refused to do so, that section of the Harris-Rosenthal paper is nowhere cited in the NRC report.

[4] Because Honorton and his colleagues have complied with the Hyman-Honorton specification that experimental reports be sufficiently complete to permit others to reconstruct the investigator's procedures, readers who wish to know more detail than we provide here are likely to find whatever they need in the archival publication of these studies in the *Journal of Parapsychology* (Honorton et al., 1990).

the TV monitor after each clip was shown. The receiver was permitted to see each clip and to change the ratings repeatedly until he or she was satisfied. The computer then wrote these and other data from the session into a file on a floppy disk. At that point, the sender moved to the receiver's chamber and revealed the identity of the target to both the receiver and the experimenter. Note that the experimenter did not even know the identity of the four-clip judging set until it was displayed to the receiver for judging.

Randomization The random selection of the target and sequencing of the judging set were controlled by a noise-based random number generator interfaced to the computer. Extensive testing confirmed that the generator was providing a uniform distribution of values throughout the full target range (1–160). Tests on the actual frequencies observed during the experiments confirmed that targets were, on average, selected uniformly from among the 4 clips within each judging set and that the 4 judging sequences used were uniformly distributed across sessions.

Additional control features The receiver's and sender's rooms were sound-isolated, electrically shielded chambers with single-door access that could be continuously monitored by the experimenter. There was two-way intercom communication between the experimenter and the receiver but only one-way communication into the sender's room; thus, neither the experimenter nor the receiver could monitor events inside the sender's room. The archival record for each session includes an audiotape containing the receiver's mentation during the ganzfeld period and all verbal exchanges between the experimenter and the receiver throughout the experiment.

The automated ganzfeld protocol has been examined by several dozen parapsychologists and behavioral researchers from other fields, including well-known critics of parapsychology. Many have participated as subjects or observers. All have expressed satisfaction with the handling of security issues and controls.

Parapsychologists have often been urged to employ magicians as consultants to ensure that the experimental protocols are not vulnerable either to inadvertent sensory leakage or to deliberate cheating. Two "mentalists," magicians who specialize in the simulation of psi, have examined the autoganzfeld system and protocol. Ford Kross, a professional mentalist and officer of the mentalist's professional organization, the Psychic Entertainers Association, provided the following written statement "In my professional capacity as a mentalist, I have reviewed Psychophysical Research Laboratories' automated ganzfeld system and found it to provide excellent security against deception by subjects" (personal communication, May, 1989).

Daryl J. Bem has also performed as a mentalist for many years and is a member of the Psychic Entertainers Association. As mentioned in the author note, this article had its origins in a 1983 visit he made to Honorton's laboratory, where he was asked to critically examine the research protocol from the perspective of a mentalist, a research psychologist, and a subject. Needless to say, this article would not exist if he did not concur with Ford Kross's assessment of the security procedures.

Experimental Studies

Altogether, 100 men and 140 women participated as receivers in 354 sessions during the research program.[5] The participants ranged in age from 17 to 74 years (M = 37.3, SD = 11.8), with a mean formal education of 15.6 years (SD = 2.0). Eight separate experimenters, including Honorton, conducted the studies.

The experimental program included three pilot and eight formal studies. Five of the formal studies used novice (first-time) participants who served as the receiver in one session each. The remaining three formal studies used experienced participants.

Pilot studies. Sample sizes were not preset in the three pilot studies. Study 1 comprised 22 sessions and was conducted during the initial development and testing of the autoganzfeld system. Study 2 comprised 9 sessions testing a procedure in which the experimenter, rather than the receiver, served as the judge at the end of the session. Study 3 comprised 35 sessions and served as practice for participants who had completed the allotted number of sessions in the ongoing formal studies but who wanted additional ganzfeld experience. This study also included several demonstration sessions when TV film crews were present.

Novice studies Studies 101–104 were each designed to test 50 participants who had had no prior ganzfeld experience; each participant served as the receiver in a single ganzfeld session. Study 104 included 16 of 20 students recruited from the Juilliard School in New York City to test an artistically gifted sample. Study 105 was initiated to accommodate the overflow of participants who had been recruited for Study 104, including the 4 remaining Juilliard students. The sample size for this study was set to 25, but only 6 sessions had been completed when the laboratory closed. For purposes of exposition, we divided the 56 sessions from Studies 104 and 105 into two parts: Study 104/105(a) comprises the 36 non-Juilliard participants, and Study 104/105(b) comprises the 20 Juilliard students.

Study 201 This study was designed to retest the most promising participants from the previous studies. The number of trials was set to 20, but only 7 sessions with 3 participants had been completed when the laboratory closed.

Study 301. This study was designed to compare static and dynamic targets. The sample size was set to 50 sessions. Twenty-five experienced participants each served as the receiver in 2 sessions. Unknown to the participants, the computer control program was modified to ensure that they would each have 1 session with a static target and 1 session with a dynamic target.

Study 302 This study was designed to examine a dynamic target set that had yielded a particularly high hit rate in the previous studies. The study involved experienced participants who had had no prior experience with this particular target set and who were unaware that only one target set was being sampled. Each served as the receiver in a single session. The design called for the study to continue until 15 sessions were completed with each of the targets, but only 25 sessions had been completed when the laboratory closed.

The 11 studies just described comprise all sessions conducted during the 6.5 years of the program. There is no "file drawer" of unreported sessions.

Results

Overall hit rate. As in the earlier meta-analysis, receivers' ratings were analyzed by tallying the proportion of hits achieved and calculating the exact binomial probability for the observed number of hits compared with the chance expectation of .25. As noted earlier, 240 participants contributed 354 sessions. For reasons discussed later, Study 302 is analyzed separately, reducing the number of sessions in the primary analysis to 329.

As Table 1 shows, there were 106 hits in the 329 sessions, a hit rate of 32% (z = 2.89, p = .002, one-tailed), with a 95% confidence interval from 30% to 35%. This corresponds to an effect size (π) of .59, with a 95% confidence interval from .53 to .64.

Table 1 also shows that when Studies 104 and 105 are combined and re-divided into Studies 104/105(a) and 104/105(b), 9

[5] A recent review of the original computer files uncovered a duplicate record in the autoganzfeld database. This has now been eliminated, reducing by one the number of subjects and sessions. As a result, some of the numbers presented in this article differ slightly from those in Honorton et al. (1990).

Table 1
Outcome by Study

Study	Study/subject description	N subjects	N trials	N hits	% hits	Effect size π	z
1	Pilot	19	22	8	36	.62	0.99
2	Pilot	4	9	3	33	.60	0.25
3	Pilot	24	35	10	29	.55	0.32
101	Novice	50	50	12	24	.47	−0.30
102	Novice	50	50	18	36	.63	1.60
103	Novice	50	50	15	30	.55	0.67
104/105(a)	Novice	36	36	12	33	.60	0.97
104/105(b)	Juilliard sample	20	20	10	50	.75	2.20
201	Experienced	3	7	3	43	.69	0.69
301	Experienced	25	50	15	30	.56	0.67
302	Experienced	25	25	16	54[a]	.78[a]	3.04[a]
Overall (Studies 1–301)		240	329	106	32	.59	2.89

Note. All z scores are based on the exact binomial probability, with $p = .25$ and $q = .75$.
[a] Adjusted for response bias; the hit rate actually observed was 64%.

of the 10 studies yield positive effect sizes, with a mean effect size (π) of .61, $t(9) = 4.44$, $p = .0008$, one-tailed. This effect size is equivalent to a four-alternative hit rate of 34%. Alternatively, if Studies 104 and 105 are retained as separate studies, 9 of the 10 studies again yield positive effect sizes, with a mean effect size (π) of .62, $t(9) = 3.73$, $p = .002$, one-tailed. This effect size is equivalent to a four-alternative hit rate of 35% and is identical to that found across the 28 studies of the earlier meta-analysis.[6]

Considered together, sessions with novice participants (Studies 101–105) yielded a statistically significant hit rate of 32.5% ($p = .009$), which is not significantly different from the 31.6% hit rate achieved by experienced participants in Studies 201 and 301. And, finally, each of the eight experimenters also achieved a positive effect size, with a mean π of .60, $t(7) = 3.44$, $p = .005$, one-tailed.

The Juilliard sample. There are several reports in the literature of a relationship between creativity or artistic ability and psi performance (Schmeidler, 1988). To explore this possibility in the ganzfeld setting, 10 male and 10 female undergraduates were recruited from the Juilliard School. Of these, 8 were music students, 10 were drama students, and 2 were dance students. Each served as the receiver in a single session in Study 104 or 105. As shown in Table 1, these students achieved a hit rate of 50% ($p = .014$), one of the five highest hit rates ever reported for a single sample in a ganzfeld study. The musicians were particularly successful: 6 of the 8 (75%) successfully identified their targets ($p = .004$; further details about this sample and their ganzfeld performance were reported in Schlitz & Honorton, 1992).

Study size and effect size. There is a significant negative correlation across the 10 studies listed in Table 1 between the number of sessions included in a study and the study's effect size (π), $r = -.64$, $t(8) = 2.36$, $p < .05$, two-tailed. This is reminiscent of Hyman's discovery that the smaller studies in the original ganzfeld database were disproportionately likely to report statistically significant results. He interpreted this finding as evidence for a bias against the reporting of small studies that fail to achieve significant results. A similar interpretation cannot be applied to the autoganzfeld studies, however, because there are no unreported sessions.

One reviewer of this article suggested that the negative correlation might reflect a decline effect in which earlier sessions of a study are more successful than later sessions. If there were such an effect, then studies with fewer sessions would show larger effect sizes because they would end before the decline could set in. To check this possibility, we computed point-biserial correlations between hits (1) or misses (0) and the session number within each of the 10 studies. All of the correlations hovered around zero; six were positive, four were negative, and the overall mean was .01.

An inspection of Table 1 reveals that the negative correlation derives primarily from the two studies with the largest effect sizes: the 20 sessions with the Juilliard students and the 7 sessions of Study 201, the study specifically designed to retest the most promising participants from the previous studies. Accordingly, it seems likely that the larger effect sizes of these two studies—and hence the significant negative correlation between the number of sessions and the effect size—reflect genuine performance differences between these two small, highly selected samples and other autoganzfeld participants.

Study 302. All of the studies except Study 302 randomly sampled from a pool of 160 static and dynamic targets. Study 302 sampled from a single, dynamic target set that had yielded a particularly high hit rate in the previous studies. The four film clips in this set consisted of a scene of a tidal wave from the movie *Clash of the Titans*, a high-speed sex scene from *A Clockwork Orange*, a scene of crawling snakes from a TV documentary, and a scene from a Bugs Bunny cartoon.

[6] As noted above, the laboratory was forced to close before three of the formal studies could be completed. If we assume that the remaining trials in Studies 105 and 201 would have yielded only chance results, this would reduce the overall z for the first 10 autoganzfeld studies from 2.89 to 2.76 ($p = .003$). Thus, inclusion of the two incomplete studies does not pose an optional stopping problem. The third incomplete study, Study 302, is discussed below.

The experimental design called for this study to continue until each of the clips had served as the target 15 times. Unfortunately, the premature termination of this study at 25 sessions left an imbalance in the frequency with which each clip had served as the target. This means that the high hit rate observed (64%) could well be inflated by response biases.

As an illustration, water imagery is frequently reported by receivers in ganzfeld sessions, whereas sexual imagery is rarely reported. (Some participants probably are reluctant both to report sexual imagery and to give the highest rating to the sex-related clip.) If a video clip containing popular imagery (such as water) happens to appear as a target more frequently than a clip containing unpopular imagery (such as sex), a high hit rate might simply reflect the coincidence of those frequencies of occurrence with participants' response biases. And, as the second column of Table 2 reveals, the tidal wave clip did in fact appear more frequently as the target than did the sex clip. More generally, the second and third columns of Table 2 show that the frequency with which each film clip was ranked first closely matches the frequency with which each appeared as the target.

One can adjust for this problem by using the observed frequencies in these two columns to compute the hit rate expected if there were no psi effect. In particular, one can multiply each proportion in the second column by the corresponding proportion in the third column—yielding the joint probability that the clip was the target and that it was ranked first—and then sum across the four clips. As shown in the fourth column of Table 2, this computation yields an overall expected hit rate of 34.08%. When the observed hit rate of 64% is compared with this baseline, the effect size (h) is .61. As shown in Table 1, this is equivalent to a four-alternative hit rate of 54%, or a π value of .78, and is statistically significant ($z = 3.04$, $p = .0012$).

The psi effect can be seen even more clearly in the remaining columns of Table 2, which control for the differential popularity of the imagery in the clips by displaying how frequently each was ranked first when it was the target and how frequently it was ranked first when it was one of the control clips (decoys). As can be seen, each of the four clips was selected as the target relatively more frequently when it was the target than when it was a decoy, a difference that is significant for three of the four clips. On average, a clip was identified as the target 58% of the time when it was the target and only 14% of the time when it was a decoy.

Dynamic versus static targets. The success of Study 302 raises the question of whether dynamic targets are, in general, more effective than static targets. This possibility was also suggested by the earlier meta-analysis, which revealed that studies using multiple-image targets (View Master stereoscopic slide reels) obtained significantly higher hit rates than did studies using single-image targets. By adding motion and sound, the video clips might be thought of as high-tech versions of the View Master reels.

The 10 autoganzfeld studies that randomly sampled from both dynamic and static target pools yielded 164 sessions with dynamic targets and 165 sessions with static targets. As predicted, sessions using dynamic targets yielded significantly more hits than did sessions using static targets (37% vs. 27%; Fisher's exact $p < .04$).

Sender–receiver pairing. The earlier meta-analysis revealed that studies in which participants were free to bring in friends

Table 2
Study 302. Expected Hit Rate and Proportion of Sessions in Which Each Video Clip Was Ranked First When It Was a Target and When It Was a Decoy

Video clip	Relative frequency as target	Relative frequency of first place ranking	Expected hit rate (%)	Ranked first when target	Ranked first when decoy	Difference	Fisher's exact p
Tidal wave	.28 (7/25)	.24 (6/25)	6.72	.57 (4/7)	.11 (2/18)	.46	.032
Snakes	.12 (3/25)	.12 (3/25)	1.44	.67 (2/3)	.05 (1/22)	.62	.029
Sex scene	.16 (4/25)	.08 (2/25)	1.28	.25 (1/4)	.05 (1/21)	.20	.300
Bugs Bunny	.44 (11/25)	.56 (14/25)	24.64	.82 (9/11)	.36 (5/14)	.46	.027
Overall			34.08	.58	.14	.44	

to serve as senders produced significantly higher hit rates than studies that used only laboratory-assigned senders. As noted, however, there is no record of how many of the participants in the former studies actually did bring in friends. Whatever the case, sender–receiver pairing was not a significant correlate of psi performance in the autoganzfeld studies: The 197 sessions in which the sender and receiver were friends did not yield a significantly higher proportion of hits than did the 132 sessions in which they were not (35% vs. 29%; Fisher's exact $p = .28$).

Correlations between receiver characteristics and psi performance. Most of the autoganzfeld participants were strong believers in psi: On a 7-point scale ranging from *strong disbelief in psi* (1) to *strong belief in psi* (7), the mean was 6.2 ($SD = 1.03$); only 2 participants rated their belief in psi below the midpoint of the scale. In addition, 88% of the participants reported personal experiences suggestive of psi, and 80% had some training in meditation or other techniques involving internal focus of attention.

All of these appear to be important variables. The correlation between belief in psi and psi performance is one of the most consistent findings in the parapsychological literature (Palmer, 1978). And, within the autoganzfeld studies, successful performance of novice (first-time) participants was significantly predicted by reported personal psi experiences, involvement with meditation or other mental disciplines, and high scores on the Feeling and Perception factors of the Myers-Briggs Type Inventory (Honorton, 1992; Honorton & Schechter, 1987; Myers & McCaulley, 1985). This recipe for success has now been independently replicated in another laboratory (Broughton, Kanthamani, & Khilji, 1990).

The personality trait of extraversion is also associated with better psi performance. A meta-analysis of 60 independent studies with nearly 3,000 subjects revealed a small but reliable positive correlation between extraversion and psi performance, especially in studies that used free-response methods of the kind used in the ganzfeld experiments (Honorton, Ferrari, & Bem, 1992). Across 14 free-response studies conducted by four independent investigators, the correlation for 612 subjects was .20 ($z = 4.82$, $p = 1.5 \times 10^{-6}$). This correlation was replicated in the autoganzfeld studies, in which extraversion scores were available for 218 of the 240 subjects, $r = .18$, $t(216) = 2.67$, $p = .004$, one-tailed.

Finally, there is the strong psi performance of the Juilliard students, discussed earlier, which is consistent with other studies in the parapsychological literature suggesting a relationship between successful psi performance and creativity or artistic ability.

Discussion

Earlier in this article, we quoted from the abstract of the Hyman–Honorton (1986) communiqué: "We agree that the final verdict awaits the outcome of future experiments conducted by a broader range of investigators and according to more stringent standards" (p. 351). We believe that the "stringent standards" requirement has been met by the autoganzfeld studies. The results are statistically significant and consistent with those in the earlier database. The mean effect size is quite respectable in comparison with other controversial research areas of human performance (Harris & Rosenthal, 1988a). And there are reliable relationships between successful psi performance and conceptually relevant experimental and subject variables, relationships that also replicate previous findings. Hyman (1991) has also commented on the autoganzfeld studies: "Honorton's experiments have produced intriguing results. If. . . independent laboratories can produce similar results with the same relationships and with the same attention to rigorous methodology, then parapsychology may indeed have finally captured its elusive quarry" (p. 392).

Issues of Replication

As Hyman's comment implies, the autoganzfeld studies by themselves cannot satisfy the requirement that replications be conducted by a "broader range of investigators." Accordingly, we hope the findings reported here will be sufficiently provocative to prompt others to try replicating the psi ganzfeld effect.

We believe that it is essential, however, that future studies comply with the methodological, statistical, and reporting standards set forth in the joint communiqué and achieved by the autoganzfeld studies. It is not necessary for studies to be as automated or as heavily instrumented as the autoganzfeld studies to satisfy the methodological guidelines, but they are still likely to be labor intensive and potentially expensive.[7]

Statistical Power and Replication

Would-be replicators also need to be reminded of the power requirements for replicating small effects. Although many academic psychologists do not believe in psi, many apparently do believe in miracles when it comes to replication. Tversky and Kahneman (1971) posed the following problem to their colleagues at meetings of the Mathematical Psychology Group and the American Psychological Association:

> Suppose you have run an experiment on 20 subjects and have obtained a significant result which confirms your theory ($z = 2.23$, $p < .05$, two-tailed). You now have cause to run an additional group of 10 subjects. What do you think the probability is that the results will be significant, by a one-tailed test, separately for this group? (p. 105)

The median estimate was .85, with 9 of 10 respondents providing an estimate greater than .60. The correct answer is approximately .48.

As Rosenthal (1990) has warned: "Given the levels of statistical power at which we normally operate, we have no right to expect the proportion of significant results that we typically do expect, even if in nature there is a very real and very important effect" (p. 16). In this regard, it is again instructive to consider the medical study that revealed a highly significant effect of aspirin on the incidence of heart attacks. The study monitored more than 22,000 subjects. Had the investigators monitored 3,000 subjects, they would have had less than an even chance of

[7] As the closing of the autoganzfeld laboratory exemplifies, it is also difficult to obtain funding for psi research. The traditional, peer-refereed sources of funding familiar to psychologists have almost never funded proposals for psi research. The widespread skepticism of psychologists toward psi is almost certainly a contributing factor.

finding a conventionally significant effect. Such is life with small effect sizes.

Given its larger effect size, the prospects for successfully replicating the psi ganzfeld effect are not quite so daunting, but they are probably still grimmer than intuition would suggest. If the true hit rate is in fact about 34% when 25% is expected by chance, then an experiment with 30 trials (the mean for the 28 studies in the original meta-analysis) has only about 1 chance in 6 of finding an effect significant at the .05 level with a one-tailed test. A 50-trial experiment boosts that chance to about 1 in 3. One must escalate to 100 trials to come close to the break-even point, at which one has a 50–50 chance of finding a statistically significant effect (Utts, 1986). (Recall that only 2 of the 11 autoganzfeld studies yielded results that were individually significant at the conventional .05 level.) Those who require that a psi effect be statistically significant every time before they will seriously entertain the possibility that an effect really exists know not what they ask.

Significance Versus Effect Size

The preceding discussion is unduly pessimistic, however, because it perpetuates the tradition of worshipping the significance level. Regular readers of this journal are likely to be familiar with recent arguments imploring behavioral scientists to overcome their slavish dependence on the significance level as the ultimate measure of virtue and instead to focus more of their attention on effect sizes: "Surely, God loves the .06 nearly as much as the .05" (Rosnow & Rosenthal, 1989, p. 1277). Accordingly, we suggest that achieving a respectable effect size with a methodologically tight ganzfeld study would be a perfectly welcome contribution to the replication effort, no matter how untenurable the p level renders the investigator.

Career consequences aside, this suggestion may seem quite counterintuitive. Again, Tversky and Kahneman (1971) have provided an elegant demonstration. They asked several of their colleagues to consider an investigator who runs 15 subjects and obtains a significant t value of 2.46. Another investigator attempts to duplicate the procedure with the same number of subjects and obtains a result in the same direction but with a nonsignificant value of t. Tversky and Kahneman then asked their colleagues to indicate the highest level of t in the replication study they would describe as a failure to replicate. The majority of their colleagues regarded $t = 1.70$ as a failure to replicate. But if the data from two such studies ($t = 2.46$ and $t = 1.70$) were pooled, the t for the combined data would be about 3.00 (assuming equal variances):

> Thus, we are faced with a paradoxical state of affairs, in which the same data that would increase our confidence in the finding when viewed as part of the original study, shake our confidence when viewed as an independent study. (Tversky & Kahneman, 1971, p. 108)

Such is the iron grip of the arbitrary .05. Pooling the data, of course, is what meta-analysis is all about. Accordingly, we suggest that two or more laboratories could collaborate in a ganzfeld replication effort by conducting independent studies and then pooling them in meta-analytic fashion, what one might call real-time meta-analysis. (Each investigator could then claim the pooled p level for his or her own curriculum vitae.)

Maximizing Effect Size

Rather than buying or borrowing larger sample sizes, those who seek to replicate the psi ganzfeld effect might find it more intellectually satisfying to attempt to maximize the effect size by attending to the variables associated with successful outcomes. Thus, researchers who wish to enhance the chances of successful replication should use dynamic rather than static targets. Similarly, we advise using participants with the characteristics we have reported to be correlated with successful psi performance. Random college sophomores enrolled in introductory psychology do not constitute the optimal subject pool.

Finally, we urge ganzfeld researchers to read carefully the detailed description of the warm social ambiance that Honorton et al. (1990) sought to create in the autoganzfeld laboratory. We believe that the social climate created in psi experiments is a critical determinant of their success or failure.

The Problem of "Other" Variables

This caveat about the social climate of the ganzfeld experiment prompted one reviewer of this article to worry that this provided "an escape clause" that weakens the falsifiability of the psi hypothesis: "Until Bem and Honorton can provide operational criteria for creating a warm social ambiance, the failure of an experiment with otherwise adequate power can always be dismissed as due to a lack of warmth."

Alas, it is true; we devoutly wish it were otherwise. But the operation of unknown variables in moderating the success of replications is a fact of life in all of the sciences. Consider, for example, an earlier article in this journal by Spence (1964). He reviewed studies testing the straightforward derivation from Hullian learning theory that high-anxiety subjects should condition more strongly than low-anxiety subjects. This hypothesis was confirmed 94% of the time in Spence's own laboratory at the University of Iowa but only 63% of the time in laboratories at other universities. In fact, Kimble and his associates at Duke University and the University of North Carolina obtained results in the opposite direction in two of three experiments.

In searching for a post hoc explanation, Spence (1964) noted that "a deliberate attempt was made in the Iowa studies to provide conditions in the laboratory that might elicit some degree of emotionality. Thus, the experimenter was instructed to be impersonal and quite formal . . . and did not try to put [subjects] at ease or allay any expressed fears" (pp. 135–136). Moreover, he pointed out, his subjects sat in a dental chair, whereas Kimble's subjects sat in a secretarial chair. Spence even considered "the possibility that cultural backgrounds of southern and northern students may lead to a difference in the manner in which they respond to the different items in the [Manifest Anxiety] scale" (p. 136). If this was the state of affairs in an area of research as well established as classical conditioning, then the suggestion that the social climate of the psi laboratory might affect the outcome of ganzfeld experiments in ways not yet completely understood should not be dismissed as a devious attempt to provide an escape clause in case of replication failure.

The best the original researchers can do is to communicate as complete a knowledge of the experimental conditions as possible in an attempt to anticipate some of the relevant moderating

variables. Ideally, this might include direct training by the original researchers or videotapes of actual sessions. Lacking these, however, the detailed description of the autoganzfeld procedures provided by Honorton et al. (1990) comes as close as current knowledge permits in providing for other researchers the "operational criteria for creating a warm social ambiance."

Theoretical Considerations

Up to this point, we have confined our discussion to strictly empirical matters. We are sympathetic to the view that one should establish the existence of a phenomenon, anomalous or not, before attempting to explain it. So let us suppose for the moment that we have a genuine anomaly of information transfer here. How can it be understood or explained?

The Psychology of Psi

In attempting to understand psi, parapsychologists have typically begun with the working assumption that, whatever its underlying mechanisms, it should behave like other, more familiar psychological phenomena. In particular, they typically assume that target information behaves like an external sensory stimulus that is encoded, processed, and experienced in familiar information-processing ways. Similarly, individual psi performances should covary with experimental and subject variables in psychologically sensible ways. These assumptions are embodied in the model of psi that motivated the ganzfeld studies in the first place.

The ganzfeld procedure. As noted in the introduction, the ganzfeld procedure was designed to test a model in which psi-mediated information is conceptualized as a weak signal that is normally masked by internal somatic and external sensory "noise." Accordingly, any technique that raises the signal-to-noise ratio should enhance a person's ability to detect psi-mediated information. This noise-reduction model of psi organizes a large and diverse body of experimental results, particularly those demonstrating the psi-conducive properties of altered states of consciousness such as meditation, hypnosis, dreaming, and, of course, the ganzfeld itself (Rao & Palmer, 1987).

Alternative theories propose that the ganzfeld (and altered states) may be psi conducive because it lowers resistance to accepting alien imagery, diminishes rational or contextual constraints on the encoding or reporting of information, stimulates more divergent thinking, or even just serves as a placebo-like ritual that participants perceive as being psi conducive (Stanford, 1987). At this point, there are no data that would permit one to choose among these alternatives, and the noise-reduction model remains the most widely accepted.

The target. There are also a number of plausible hypotheses that attempt to account for the superiority of dynamic targets over static targets: Dynamic targets contain more information, involve more sensory modalities, evoke more of the receiver's internal schemata, are more lifelike, have a narrative structure, are more emotionally evocative, and are "richer" in other, unspecified ways. Several psi researchers have attempted to go beyond the simple dynamic–static dichotomy to more refined or theory-based definitions of a good target. Although these efforts have involved examining both psychological and physical properties of targets, there is as yet not much progress to report (Delanoy, 1990).

The receiver. Some of the subject characteristics associated with good psi performance also appear to have psychologically straightforward explanations. For example, garden-variety motivational explanations seem sufficient to account for the relatively consistent finding that those who believe in psi perform significantly better than those who do not. (Less straightforward, however, would be an explanation for the frequent finding that nonbelievers actually perform significantly worse than chance [Broughton, 1991, p. 109].)

The superior psi performance of creative or artistically gifted individuals—such as the Juilliard students—may reflect individual differences that parallel some of the hypothesized effects of the ganzfeld mentioned earlier: Artistically gifted individuals may be more receptive to alien imagery, be better able to transcend rational or contextual constraints on the encoding or reporting of information, or be more divergent in their thinking. It has also been suggested that both artistic and psi abilities might be rooted in superior right-brain functioning.

The observed relationship between extraversion and psi performance has been of theoretical interest for many years. Eysenck (1966) reasoned that extraverts should perform well in psi tasks because they are easily bored and respond favorably to novel stimuli. In a setting such as the ganzfeld, extraverts may become "stimulus starved" and thus may be highly sensitive to any stimulation, including weak incoming psi information. In contrast, introverts would be more inclined to entertain themselves with their own thoughts and thus continue to mask psi information despite the diminished sensory input. Eysenck also speculated that psi might be a primitive form of perception antedating cortical developments in the course of evolution, and, hence, cortical arousal might suppress psi functioning. Because extraverts have a lower level of cortical arousal than introverts, they should perform better in psi tasks (the evolutionary biology of psi has also been discussed by Broughton, 1991, pp. 347–352).

But there are more mundane possibilities. Extraverts might perform better than introverts simply because they are more relaxed and comfortable in the social setting of the typical psi experiment (e.g., the "warm social ambiance" of the autoganzfeld studies). This interpretation is strengthened by the observation that introverts outperformed extraverts in a study in which subjects had no contact with an experimenter but worked alone at home with materials they received in the mail (Schmidt & Schlitz, 1989). To help decide among these interpretations, ganzfeld experimenters have begun to use the extraversion scale of the NEO Personality Inventory (Costa & McCrae, 1992), which assesses six different facets of the extraversion–introversion factor.

The sender. In contrast to this information about the receiver in psi experiments, virtually nothing is known about the characteristics of a good sender or about the effects of the sender's relationship with the receiver. As has been shown, the initial suggestion from the meta-analysis of the original ganzfeld database that psi performance might be enhanced when the sender and receiver are friends was not replicated at a statistically significant level in the autoganzfeld studies.

A number of parapsychologists have entertained the more

radical hypothesis that the sender may not even be a necessary element in the psi process. In the terminology of parapsychology, the sender–receiver procedure tests for the existence of *telepathy,* anomalous communication between two individuals; however, if the receiver is somehow picking up the information from the target itself, it would be termed clairvoyance, and the presence of the sender would be irrelevant (except for possible psychological reasons, such as expectation effects).

At the time of his death, Honorton was planning a series of autoganzfeld studies that would systematically compare sender and no-sender conditions while keeping both the receiver and the experimenter blind to the condition of the ongoing session. In preparation, he conducted a meta-analytic review of ganzfeld studies that used no sender. He found 12 studies, with a median of 33.5 sessions, conducted by seven investigators. The overall effect size (π) was .56, which corresponds to a four-alternative hit rate of 29%. But this effect size does not reach statistical significance (Stouffer $z = 1.31$, $p = .095$). So far, then, there is no firm evidence for psi in the ganzfeld in the absence of a sender. (There are, however, nonganzfeld studies in the literature that do report significant evidence for clairvoyance, including a classic card-guessing experiment conducted by J. B. Rhine and Pratt [1954].)

The Physics of Psi

The psychological level of theorizing just discussed does not, of course, address the conundrum that makes psi phenomena anomalous in the first place: their presumed incompatibility with our current conceptual model of physical reality. Parapsychologists differ widely from one another in their taste for theorizing at this level, but several whose training lies in physics or engineering have proposed physical (or biophysical) theories of psi phenomena (an extensive review of theoretical parapsychology was provided by Stokes, 1987). Only some of these theories would force a radical revision in our current conception of physical reality.

Those who follow contemporary debates in modern physics, however, will be aware that several phenomena predicted by quantum theory and confirmed by experiment are themselves incompatible with our current conceptual model of physical reality. Of these, it is the 1982 empirical confirmation of Bell's theorem that has created the most excitement and controversy among philosophers and the few physicists who are willing to speculate on such matters (Cushing & McMullin, 1989; Herbert, 1987). In brief, Bell's theorem states that any model of reality that is compatible with quantum mechanisms must be *nonlocal*: It must allow for the possibility that the results of observations at two arbitrarily distant locations can be correlated in ways that are incompatible with any physically permissible causal mechanism.

Several possible models of reality that incorporate nonlocality have been proposed by both philosophers and physicists. Some of these models clearly rule out psi-like information transfer, others permit it, and some actually require it. Thus, at a grander level of theorizing, some parapsychologists believe that one of the more radical models of reality compatible with both quantum mechanics and psi will eventually come to be accepted. If and when that occurs, psi phenomena would cease to be anomalous.

But we have learned that all such talk provokes most of our colleagues in psychology and in physics to roll their eyes and gnash their teeth. So let's just leave it at that.

Skepticism Revisited

More generally, we have learned that our colleagues' tolerance for *any* kind of theorizing about psi is strongly determined by the degree to which they have been convinced by the data that psi has been demonstrated. We have further learned that their diverse reactions to the data themselves are strongly determined by their a priori beliefs about and attitudes toward a number of quite general issues, some scientific, some not. In fact, several statisticians believe that the traditional hypothesis-testing methods used in the behavioral sciences should be abandoned in favor of Bayesian analyses, which take into account a person's a priori beliefs about the phenomenon under investigation (e.g., Bayarri & Berger, 1991; Dawson, 1991).

In the final analysis, however, we suspect that both one's Bayesian a prioris and one's reactions to the data are ultimately determined by whether one was more severely punished in childhood for Type I or Type II errors.

References

Atkinson, R., Atkinson, R. C., Smith, E. E., & Bem, D. J. (1990). *Introduction to psychology* (10th ed.). San Diego, CA: Harcourt Brace Jovanovich.

Atkinson, R., Atkinson, R. C., Smith, E. E., & Bem, D. J. (1993). *Introduction to psychology* (11th ed.). San Diego, CA: Harcourt Brace Jovanovich.

Avant, L. L. (1965). Vision in the ganzfeld. *Psychological Bulletin, 64,* 246–258.

Bayarri, M. J., & Berger, J. (1991). Comment. *Statistical Science, 6,* 379–382.

Blackmore, S. (1980). The extent of selective reporting of ESP ganzfeld studies. *European Journal of Parapsychology, 3,* 213–219.

Bozarth, J. D., & Roberts, R. R. (1972). Signifying significant significance. *American Psychologist, 27,* 774–775.

Braud, W. G., Wood, R., & Braud, L. W. (1975). Free-response GESP performance during an experimental hypnagogic state induced by visual and acoustic ganzfeld techniques. A replication and extension. *Journal of the American Society for Psychical Research, 69,* 105–113.

Broughton, R. S. (1991). *Parapsychology: The controversial science.* New York: Ballantine Books.

Broughton, R. S., Kanthamani, H., & Khilji, A. (1990). Assessing the PRL success model on an independent ganzfeld data base. In L. Henkel & J. Palmer (Eds.), *Research in parapsychology 1989* (pp. 32–35). Metuchen, NJ: Scarecrow Press.

Child, I. L. (1985). Psychology and anomalous observations: The question of ESP in dreams. *American Psychologist, 40,* 1219–1230.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1,* 98–101.

Costa, P. T. Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional manual.* Odessa, FL: Psychological Assessment Resources.

Cushing, J. T., & McMullin, E. (Eds.). (1989). *Philosophical conse-*

quences of quantum theory: Reflections on Bell's theorem. Notre Dame, IN: University of Notre Dame Press.

Dawson, R. (1991). Comment. *Statistical Science, 6,* 382–385.

Delanoy, D. L. (1990). Approaches to the target: A time for reevaluation. In L. A. Henkel & J. Palmer (Eds.), *Research in parapsychology 1989* (pp. 89–92). Metuchen, NJ: Scarecrow Press.

Dingwall, E. J. (Ed.). (1968). *Abnormal hypnotic phenomena* (4 vols.). London: Churchill.

Druckman, D., & Swets, J. A. (Eds.). (1988). *Enhancing human performance: Issues, theories, and techniques.* Washington, DC: National Academy Press.

Eysenck, H. J. (1966). Personality and extra-sensory perception. *Journal of the Society for Psychical Research, 44,* 55–71.

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life.* New York: Free Press.

Green, C. E. (1960). Analysis of spontaneous cases. *Proceedings of the Society for Psychical Research, 53,* 97–161.

Harris, M. J., & Rosenthal, R. (1988a). *Human performance research: An overview.* Washington, DC: National Academy Press.

Harris, M. J., & Rosenthal, R. (1988b). *Postscript to "Human performance research: An overview."* Washington, DC: National Academy Press.

Herbert, N. (1987). *Quantum reality: Beyond the new physics.* Garden City, NY: Anchor Books.

Honorton, C. (1969). Relationship between EEG alpha activity and ESP card-guessing performance. *Journal of the American Society for Psychical Research, 63,* 365–374.

Honorton, C. (1977). Psi and internal attention states. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 435–472). New York: Van Nostrand Reinhold.

Honorton, C. (1979). Methodological issues in free-response experiments. *Journal of the American Society for Psychical Research, 73,* 381–394.

Honorton, C. (1985). Meta-analysis of psi ganzfeld research: A response to Hyman. *Journal of Parapsychology, 49,* 51–91.

Honorton, C. (1992, August). *The ganzfeld novice: Four predictors of initial ESP performance.* Paper presented at the 35th annual convention of the Parapsychological Association, Las Vegas, NV.

Honorton, C., Berger, R. E., Varvoglis, M. P., Quant, M., Derr, P., Schechter, E. I., & Ferrari, D. C. (1990). Psi communication in the ganzfeld: Experiments with an automated testing system and a comparison with a meta-analysis of earlier studies. *Journal of Parapsychology, 54,* 99–139.

Honorton, C., Ferrari, D. C., & Bem, D. J. (1992). Extraversion and ESP performance: Meta-analysis and a new confirmation. In L. A. Henkel & G. R. Schmeidler (Eds.), *Research in parapsychology 1990* (pp. 35–38). Metuchen, NJ: Scarecrow Press.

Honorton, C., & Harper, S. (1974). Psi-mediated imagery and ideation in an experimental procedure for regulating perceptual input. *Journal of the American Society for Psychical Research, 68,* 156–168.

Honorton, C., & Schechter, E. I. (1987). Ganzfeld target retrieval with an automated testing system: A model for initial ganzfeld success. In D. B. Weiner & R. D. Nelson (Eds.), *Research in parapsychology 1986* (pp. 36–39). Metuchen, NJ: Scarecrow Press.

Hyman, R. (1985). The ganzfeld psi experiment: A critical appraisal. *Journal of Parapsychology, 49,* 3–49.

Hyman, R. (1991). Comment. *Statistical Science, 6,* 389–392.

Hyman, R., & Honorton, C. (1986). A joint communiqué: The psi ganzfeld controversy. *Journal of Parapsychology, 50,* 351–364.

Kennedy, J. E. (1979). Methodological problems in free-response ESP experiments. *Journal of the American Society for Psychical Research, 73,* 1–15.

Metzger, W. (1930). Optische Untersuchungen am Ganzfeld: II. Zur phanomenologie des homogenen Ganzfelds [Optical investigation of the Ganzfeld: II Toward the phenomenology of the homogeneous Ganzfeld]. *Psychologische Forschung, 13,* 6–29.

Morris, R. L. (1991). Comment. *Statistical Science, 6,* 393–395.

Myers, I. B. & McCaulley, M. H. (1985). *Manual: A guide to the development and use of the Myers-Briggs Type Indicator.* Palo Alto, CA: Consulting Psychologists Press.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice Hall.

Palmer, J. (1978). Extrasensory perception: Research findings. In S. Krippner (Ed.), *Advances in parapsychological research* (Vol. 2, pp. 59–243). New York: Plenum Press.

Palmer, J. A., Honorton, C., & Utts, J. (1989). Reply to the National Research Council Study on Parapsychology. *Journal of the American Society for Psychical Research, 83,* 31–49.

Parker, A. (1975). Some findings relevant to the change in state hypothesis. In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.), *Research in parapsychology, 1974* (pp. 40–42). Metuchen, NJ: Scarecrow Press.

Parker, A. (1978). A holistic methodology in psi research. *Parapsychology Review, 9,* 1–6.

Prasad, J., & Stevenson, I. (1968). A survey of spontaneous psychical experiences in school children of Uttar Pradesh, India. *International Journal of Parapsychology, 10,* 241–261.

Rao, K. R., & Palmer, J. (1987). The anomaly called psi: Recent research and criticism. *Behavioral and Brain Sciences, 10,* 539–551.

Rhine, J. B., & Pratt, J. G. (1954). A review of the Pearce-Pratt distance series of ESP tests. *Journal of Parapsychology, 18,* 165–177.

Rhine, L. E. (1962). Psychological processes in ESP experiences. I. Waking experiences. *Journal of Parapsychology, 26,* 88–111.

Roig, M., Icochea, H., & Cuzzucoli, A. (1991). Coverage of parapsychology in introductory psychology textbooks. *Teaching of Psychology, 18,* 157–160.

Rosenthal, R. (1978). Combining results of independent studies. *Psychological Bulletin, 85,* 185–193.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86,* 638–641.

Rosenthal, R. (1990). Replication in behavioral research. *Journal of Social Behavior and Personality, 5,* 1–30.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Newbury Park, CA: Sage.

Rosenthal, R., & Rubin, D. B. (1989). Effect size estimation for one-sample multiple-choice-type data: Design, analysis, and meta-analysis. *Psychological Bulletin, 106,* 332–337.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44,* 1276–1284.

Sannwald, G. (1959). Statistische untersuchungen an Spontanphänomene [Statistical investigation of spontaneous phenomena]. *Zeitschrift für Parapsychologie and Grenzgebiete der Psychologie, 3,* 59–71.

Saunders, D. R. (1985). On Hyman's factor analyses. *Journal of Parapsychology, 49,* 86–88.

Schechter, E. I. (1984). Hypnotic induction vs. control conditions: Illustrating an approach to the evaluation of replicability in parapsychology. *Journal of the American Society for Psychical Research, 78,* 1–27.

Schlitz, M. J., & Honorton, C. (1992). Ganzfeld psi performance within an artistically gifted population. *Journal of the American Society for Psychical Research, 86,* 83–98.

Schmeidler, G. R. (1988). *Parapsychology and psychology: Matches and mismatches.* Jefferson, NC: McFarland.

Schmidt, H., & Schlitz, M. J. (1989). A large scale pilot PK experiment with prerecorded random events. In L. A. Henkel & R. E. Berger (Eds.), *Research in parapsychology 1988* (pp. 6–10). Metuchen, NJ: Scarecrow Press.

Spence, K. W. (1964). Anxiety (drive) level and performance in eyelid conditioning. *Psychological Bulletin, 61,* 129–139.

Stanford, R. G. (1987). Ganzfeld and hypnotic-induction procedures in ESP research: Toward understanding their success. In S. Krippner (Ed.), *Advances in parapsychological research* (Vol. 5, pp. 39–76). Jefferson, NC: McFarland.

Steering Committee of the Physicians' Health Study Research Group. (1988). Preliminary report: Findings from the aspirin component of the ongoing Physicians' Health Study. *New England Journal of Medicine, 318,* 262–264.

Sterling, T. C. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association, 54,* 30–34.

Stokes, D. M. (1987). Theoretical parapsychology. In S. Krippner (Ed.), *Advances in parapsychological research* (Vol. 5, pp. 77–189). Jefferson, NC: McFarland.

Swets, J. A., & Bjork, R. A. (1990). Enhancing human performance: An evaluation of "new age" techniques considered by the U.S. Army. *Psychological Science, 1,* 85–96.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 2,* 105–110.

Ullman, M., Krippner, S., & Vaughan, A. (1973). *Dream telepathy.* New York: Macmillan.

Utts, J. (1986). The ganzfeld debate: A statistician's perspective. *Journal of Parapsychology, 50,* 393–402.

Utts, J. (1991a). Rejoinder. *Statistical Science, 6,* 396–403.

Utts, J. (1991b). Replication and meta-analysis in parapsychology. *Statistical Science, 6,* 363–378.

Wagner, M. W., & Monnet, M. (1979). Attitudes of college professors toward extrasensory perception. *Zetetic Scholar, 5,* 7–17.

Received September 28, 1992
Revision received March 10, 1993
Accepted March 14, 1993 ■

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[17]

SHOULD GANZFELD RESEARCH CONTINUE TO BE CRUCIAL IN THE SEARCH FOR A REPLICABLE PSI EFFECT? PART I. DISCUSSION PAPER AND INTRODUCTIONTO AN ELECTRONIC-MAIL DISCUSSION

By Julie Milton

ABSTRACT: A group of recent, well-controlled ganzfeld studies failed to replicate the positive findings of earlier work (Milton & Wiseman, 1999a). This presents a challenge to claims that a ganzfeld psi effect can be replicated across experimenters under methodologically stringent conditions. Because of the ganzfeld's history as a focus for proof-oriented questions, this situation has implications for parapsychology as a whole. In this paper, it is shown that replication of effect size in the recent ganzfeld studies is not demonstrated across experimenters, regardless of whether the database is updated to include recent studies or whether outcome and cumulation statistics different from those preplanned are applied. Problems with interpreting as strong evidence for psi other parapsychological meta-analyses of less clearly well-conducted studies and apparently consistent process-oriented findings are discussed. The case is made for continuing with ganzfeld research as an important focus of parapsychology's claims for replicability. It is argued that if there is a replicable ganzfeld psi effect, however, the procedures necessary to produce it have not yet been identified. It is proposed that process-oriented work be directed to the goal of identifying which studies should be able to replicate an above-chance effect, and that these studies, identified by their planned procedures before they have been conducted, should provide the basis for future tests of replication.

The organization of an international, electronic-mail discussion of these issues among 41 researchers with a special interest in ganzfeld research is described. The edited transcript of the discussion is presented in Part II.

Discussion Paper

Despite the field's long history, there is still controversy over whether the results of parapsychology experiments offer evidence for a genuine communication anomaly—*psi*. For some time, parapsychologists have

The writing of this paper and organization of the debate were generously supported by the Fundação Bial and the Society for Psychical Research. I am grateful to Hoyt Edge for moderating the debate, Gertrude Schmeidler for editing the debate material, and Paul Stevens for writing the anonymizing software and acting as systems manager for the discussion. I am indebted to the researchers who were kind enough to observe or participate in the debate and to Bob Morris for comments on an earlier draft of the discussion paper.

recognized that the evidence for psi most likely to convince fair-minded but critical scientists would be an experimental procedure that a range of experimenters could carry out that would produce reasonably replicable effects. Unless the experiment's effects could be replicated across experimenters, there would always remain fraud, error, or sensory leakage as strong alternative explanations to the psi hypothesis.

For many years, such replicability appeared to be out of reach. This perception appeared to change however, with the arrival in the 1970s of several research programs involving free-response ESP. In particular, ganzfeld ESP studies seemed especially promising. Not only did a range of experimenters appear to obtain outcomes in ganzfeld studies that were above chance, but they did so under conditions that appeared to be well-controlled and without using specially selected participants. In 1981, Ray Hyman, a psychologist skeptical of the existence of psi, wanted to conduct a critical assessment of a research program that represented parapsychology's strongest evidence. Because of claims then being made for ganzfeld research, it was an obvious choice for his attention (Hyman, 1985). Hyman (1985) meta-analyzed the 42 studies conducted since publication of the first ganzeld ESP study in 1974, finding an overall statistically significant outcome; however, he concluded that the methodological problems that he identified in the studies could account for the positive results. In response, Charles Honorton, a proponent of ganzfeld research, conducted his own meta-analysis of the database, restricting his attention to the 28 studies reporting direct hits as an outcome measure (Honorton, 1985). He also obtained a statistically significant overall outcome (see Table 1); but although he conceded that the studies contained potential methodological problems, he did not agree that the problems were sufficient to account for the overall outcome.

Rather than continue to dispute the matter, Hyman and Honorton (1986) instead jointly drew up a set of methodological guidelines for the stringent conduct of future ganzfeld studies, agreeing that the case for psi in the ganzfeld would rely on a broad range of experimenters obtaining positive results under such conditions. Meanwhile, Honorton and his research team at Princeton Research Laboratories (PRL) had begun in 1982 a series of partially automated ganzfeld studies—*autoganzfeld studies*—designed to meet Hyman's methodological concerns (Bem & Honorton, 1994; Honorton et al., 1990). Before PRL closed in 1989, eleven series were completed, obtaining a statistically significant overall outcome and a mean effect size nearly identical to that obtained in Honorton's (1985) meta-analysis of the earlier ganzfeld database (see Table 1). Replication under stringent conditions of the early ganzfeld results appeared to suggest that methodological problems were unlikely to have accounted entirely for the effects obtained in the earlier studies; however, Bem and Honorton pointed out that it still remained for their

results to be replicated by other experimenters under similarly stringent conditions.

In early 1997, Richard Wiseman and I, in an attempt to determine whether other experimenters had indeed succeeded in replicating these results under well-controlled conditions, meta-analyzed the 30 published ganzfeld studies conducted since the publication of Hyman and Honorton's methodological guidelines (Milton & Wiseman, 1999a). The studies' combined outcome was not statistically significant, and the mean effect size was near zero (see Table 1). The mean effect size in the recent studies is less than a seventeenth of that found in the PRL work, and a post hoc comparison shows that it is statistically significantly lower than the mean effect sizes of the PRL and earlier ganzfeld databases (see Table A2).

Updating our meta-analysis to include the studies (see Table A1) published to date (March 1999) since our meta-analysis was completed in February 1997 renders the overall cumulation statistically significant,[1] but fails to raise the mean effect size to even a sixth of that obtained in the PRL or earlier ganzfeld studies meta-analyzed by Hyman (1985) and Honorton (1985) (see Table 1). Moreover, the statistical significance of the updated cumulation is due not to renewed success by a range of investigators, but solely to the inclusion of an extremely successful study by Dalton (1997a) (see Table 1). Whether Dalton's study is included or not, it is clear that the effect size obtained in Honorton's autoganzfeld studies and in the earlier ganzfeld database has not replicated. Post hoc comparisons show that the updated database of recent studies, with or without the Dalton study, has a mean effect size statistically signficantly lower than those of the earlier meta-analyses (see Table A2).

The same is true if a variety of alternative outcome calculation and cumulation methods are used to analyse the recent studies rather than the ones that we preplanned and applied (Milton & Wiseman, 1997a). Since the presentation of our meta-analysis at the 1997 Parapsychological Association Annual Convention, a number of colleagues have informally pointed out that using several different methods of calculating or cumulating individual study outcomes, or introducing various criteria for excluding outliers, results in overall statistical significance of varying degrees for the database. Regardless of arguments over the post hoc and possibly selective nature of these analyses, none of them has the effect of raising the mean effect size in the new database by any meaningful amount, because of the relative insensitivity of means compared to the statistical significance of cumulations when slight changes are made in

[1] It was not possible to calculate outcomes for three of the studies (see footnotes to Table A1) but given that one of these studies (Parker & Westerlund, 1998, Serial Ganzfeld) is clearly slightly below chance and the remaining two studies are very small with only 12 trials each, it is unlikely that their results would increase the cumulated outcome of the database by a meaningful amount.

Table 1
OUTCOMES OF META-ANALYSES OF ESP GANZFELD STUDIES[a].

Meta-analysis	Number of studies	Number of trials	Stouffer *z*	*p* (1-t)	Effect size[b] Mean	Standard deviation	95% confidence interval
Honorton (1985)[c]	28	835	6.60	2.2×10^{-11}	.26	.38	.12 to .40
Bem & Honorton (1994)	11	329	3.41	.00033	.23	.24	**.09** to .37
Milton & Wiseman (1999a)	30	1198	.70	.24	**.013**	.23	**-.07** to .10
All studies 1987 to present[d]	39	1588	2.28	.011	**.038**	.26	**-.04** to .12
All studies 1987 to present excluding Dalton (1997a)[d]	38	1460	1.45	.074	**.027**	.25	**-.05** to .11

[a]The original table included some errata. The present table has been corrected, and the corrected figures are shown in bold.
[b]Effect size is $z/N^{1/2}$
[c]Here Honorton's meta-analysis solely represents the early ganzfeld database because Hyman's (1985) report does not provide the number of trials in each study needed for the calculation of $z/N^{1/2}$, the effect size used in this table.
[d]Individual study outcomes were calculated following the same procedures as in Milton and Wiseman (1999a).

the treatment of a database. For example, using Bem and Honorton's (1994) method to sum the number of direct hits obtained across studies (approximating the number of direct hits from the standard normal deviate of the study's reported outcome measure if direct hits were not reported) results in a total of 331 hits in the 1198 trials in the database. This is a statistically significant outcome, $p = .019$, one-tailed; but the effect size measured in this way is only .06.

Implications of the current situation

The current situation, then, is that the studies that appear to form the group proposed by Bem and Honorton (1994) to form a crucial test of the evidence for psi in the ganzfeld have clearly failed to show replication of an above-chance effect across experimenters; and, to date, they only show overall statistical significance if one extremely successful study is included. On the face of it, this appears to be an important replication failure because the unique history of ganzfeld research—strong claim, critical assessment, methodological guidelines, methodological refinement, initial replication—has led to it being presented to mainstream science as a critical test of the evidence for psi.

It has been almost 20 years, however, since Hyman's (1985) meta-analysis placed the focus for assessing the evidence for psi on ganzfeld research. Since that time, meta-analyses have been conducted of other parapsychological databases, including some whose main purpose has been to examine process-oriented hypotheses. The studies within them are not as well-controlled as the recent ganzfeld studies appear to be; but their highly statistically significant cumulated outcomes, their apparent resistance to explanations in terms of selective reporting, their general lack of statistically significant correlations between individual studies' quality and effect size in these databases, and the apparent replicability across experimenters of successful studies within them has led to their being presented both within and outside parapsychology as providing strong evidence that psi is a genuine communication anomaly that replicates across experimenters (e.g. Honorton & Ferrari, 1989; Radin, 1997; Radin & Ferrari, 1991; Radin & Nelson, 1989; Utts, 1991). If they do indeed constitute strong evidence, then the replication failure of the recent ganzfeld studies requires no negative reassessment of the claims for psi nor any action to continue to seek evidence for across-experimenter replication of a psi effect under stringent conditions in the ganzfeld.

TABLE 2
MEAN METHODOLOGICAL QUALITY OF STUDIES IN PARAPSYCHOLOGY META-ANALYSES EXPRESSED AS A PERCENTAGE OF THE MAXIMUM NUMBER OF QUALITY POINTS AVAILABLE.

Meta-analysis	Effect examined	Mean quality (%)
Honorton (1985)	Ganzfeld ESP	70[a]
Hyman (1985)	Ganzfeld ESP	44[b]
Honorton & Ferrari (1989)	Forced-choice precognition	41
Honorton et al. (1998)	ESP-extraversion relationship	
	Forced-choice studies	45
	Free-response studies	86
Lawrence (1993)	ESP-belief in psi relationship	46
Milton (1997)	Non-ASC free-response ESP	
	GESP studies[c]	61
	Clairvoyance studies[c]	58
	Precognition studies[c]	47
Radin & Ferrari (1991)	Dice PK	Not reported
Radin & Nelson (1989)	Micro-PK	Not reported
Stanford & Stein (1994)	ESP-Hypnosis relationship	49
Steinkamp et al. (1998)	Precognition vs clairvoyance	
	Clairvoyance studies	66
	Precognition studies	63

Note: The meta-analyses used different quality criteria, ranging from 2 to 18 safeguards being examined in each meta-analysis. The mean quality of each meta-analysis is therefore, not directly comparable with another.

[a] In this meta-analysis, Honorton assessed study quality on just two features—the availability of sensory cues from target handling and the adequacy of the target randomization method. He assigned partial credit to studies containing methodological features (the use of single rather than duplicate target sets and randomization using hand shuffling, coin-flipping or die-throwing) that have received no credit in other parapsychological meta-analyses (Honorton & Ferrari, 1989; Lawrence, 1993; Milton, 1997; etc.). This method allowed him to make a distinction between these studies and studies using less stringent or unknown methods; but for the purposes of this table, the method arguably inflates apparent study quality by a considerable amount. For example, all but one study received at least one quality point for preventing sensory cueing regardless of whether a duplicate target set was used. If quality points are assigned in a manner more consistent with the other meta-analyses, with one point for the use of duplicate judging sets and no points for manual methods of randomization, the studies obtained 46% of the maximum available quality points.

[b] Based on only 4 of Hyman's 12 flaw categories. One of the excluded categories involved assigning a flaw to studies in which it was not clear that receivers' friends were used as senders. This does not seem appropriate because it is absence of appropriate security rather than the relationship between participants that would constitute an inadequate precaution against collusion. The remaining 7 flaws concerned statistical errors and the use of multiple outcome measures without adjustment for multiple analysis. They could not have affected study outcomes in the meta-analysis because Hyman calculated outcomes using appropriate statistics and single measures and are not therefore included here.

[c] The original paper reports these percentages in terms of publication type rather than study type.

Problems in interpreting meta-analyses of studies of uncertain quality

Even if internal analyses reveal no obvious problems, there are difficulties in interpreting meta-analyses as strong evidence for a phenomenon if the studies they contain are of uncertain or low methodological quality. As can be seen in Table 2, the parapsychological databases examined so far consist of exactly such studies. The table summarizes the methodological quality observed in the major parapsychology databases meta-analyzed so far that have included individual study quality assessments. Setting aside Honorton's (1985) and Hyman's (1985) quality assessments of the early ganzfeld work, which present some problems of interpretation (see footnotes to the table), it can be seen that in fully half of the databases that reported mean study quality, studies scored on average fewer than half of the available methodological quality points. Only the 14-study free-response sub-database in Honorton, Ferrari, and Bem's (1998) meta-analysis contained studies that scored more than two-thirds of the available quality points; and it can be argued that omitted from that quality assessment were important quality criteria, such as the prespecification of sample size, the use of blind mentation transcription, the prevention of cues to judges from judging trials out of order, and so on (see Milton, 1997). Two meta-analyses did not report mean study quality at all.

The lack of evidence that these databases in general consist of high quality studies introduces the possibility that their outcomes may have been inflated or, at worst, entirely caused by methodological flaws. To be a matter for concern in parapsychology databases, the effect sizes due to methodological flaws would have to be at least as large as the observed effect sizes, and the flaws would have to be present in sufficient quantities (singly or in combination) to be relevant. There has been, however, very little empirical research to determine the effect sizes associated with the absence of the various methodological safeguards used in parapsychology (Milton & Wiseman, 1997b), and many meta-analyses do not report the frequency with which individual safeguards are not reported. It is, therefore, difficult to rule out methodological problems as an explanation for the observed results. There are, in fact, meta-analyses in which flaws likely to be associated with effect sizes not much, if any, smaller than those observed appear to be potentially prevalent. For example, it is clear that if the experimenter does not prespecify which of several possible measures (such as direct hits, ranks, etc. in free-response ESP studies) is to be used to test the null hypothesis, there is a potential to inflate study outcomes considerably, due to post hoc data selection. The effect size associated with such selection has not been calculated; however, a computer simulation by Hyman (1985) of the effects of being free to choose any of the four main outcome measures available when target ratings are

used suggests that the probability of any one of them being statistically significant with an alpha of .05 is approximately .15. In a database of 78 free-response studies (Milton, 1997), the observed probability of a study being statistically significantly above chance was .22, and 96% of studies did not report whether the choice of outcome measure was preplanned. Hyman's study is likely to provide an extreme upper limit for the action of this particular flaw because it is not probable that post hoc selection of statistically significant outcome measures happens in every study, as it did in his simulation. Nevertheless, the potential effects of not prespecifying outcome measures is clearly not trivial in comparison with the outcomes of ESP studies. Similarly, recording errors have been estimated empirically to occur on approximately 1% of trials and to be biased in favor of the observer's hypothesis on two-thirds of the trials (Rosenthal, 1978). The mean effect size in Honorton and Ferrari's (1989) database of forced-choice precognition studies is equivalent to raising a study's outcome 1% above a mean chance expectation of 50%; but the frequency with which studies reported double-blind, double-checked, or automated data recording is not reported.

In most parapsychological meta-analyses, estimates of overall study quality do not correlate statistically significantly with effect size. A number of the researchers who obtained such null correlations have concluded that methodological problems, therefore, had no meaningful influence on their databases (e.g., Honorton & Ferrari, 1989; Lawrence, 1993; Radin & Ferrari, 1991; Radin & Nelson, 1989); however, in databases that do not consist entirely or mostly of clearly well-controlled studies such as the parapsychology databases , there are many ways in which a relationship between methodological flaws and effect size could be obscured. This is a general problem in meta-analysis and not one restricted to parapsychology. Because these problems have received little attention in parapsychology (although see Hyman, 1985; Milton, 1997; Stanford & Stein, 1994), it is worth listing some of them. A selection, by no means exhaustive, is as follows:

1. The absence of safeguards for certain procedures (such as randomization or sensory-shielding procedures) might inflate effect size more than the absence of safeguards for others (such as lack of double-blind checking of data records). In an unweighted correlation of study quality and effect size, the effect of the absence of these more important safeguards might be drowned out by the other data (Stanford & Stein, 1994). In some cases, experts have been called upon to rate flaws in terms of their likely impact so that a weighted correlation can be performed between the absence of safeguards and effect size (e.g., Milton, 1997; Radin & Ferrari, 1991). Thus far, these weightings have not indicated any such relationships, but it could be argued that, given the general lack of direct empirical evidence concerning effect sizes that result

from the absence of safeguards, the experts' judgments may be wrong, regardless of how well they agree with each other.

2. It is unlikely that individual studies' methodological quality is accurately reflected by their quality coding. Most parapsychological studies, especially those conducted before the 1980s, have not been written with a future meta-analyst's quality checklist in mind; and it is often unclear from reports whether particular safeguards against sensory leakage, error, post hoc data selection, and so on have been carried out. Presented with unclear or circumstantial evidence concerning the presence of a safeguard, coders will have to make a subjective judgment according to this partial, ambiguous information, influenced by their individual expectations and assumptions about what experimenters are likely to do as a matter of standard laboratory procedure. Under these circumstances, errors in coding are very likely to arise.

3. The binary coding of methodological safeguards as either present or absent in almost all parapsychology meta-analyses to date means that studies whose use of the safeguard is unknown must be included in the "safeguard present" or "safeguard absent" group. For example, it may be assumed that studies whose reports do not address at all the issue of study size belong with studies that clearly did not prespecify, as a safeguard against optional stopping, the number of trials to be conducted (Milton & Wiseman, 1997b). However, given that at least some, but by no means all, experimenters are likely to have used the safeguard without reporting it, this will result in a group of studies that all used the safeguard being compared with a group of studies in which some used the safeguard and some did not, in an unknown proportion. If the studies that did not use the safeguard had higher effect sizes as a result, then including the studies that used but did not report the safeguard in the same category will reduce the average effect size in that group, bringing it closer to that of the group that clearly used the safeguard. Clearly, this would reduce the sensitivity of a test comparing the mean effect sizes in the two groups and could obscure a genuine relationship between effect size and methodological quality. The only parapsychological meta-analysis published so far that allowed assessors to code the presence of a safeguard in a study as unknown rather than merely present or absent found that up to 59% of studies fell into this category on certain safeguards (Steinkamp, Milton, & Morris, 1998), suggesting that the problem may be by no means trivial in other parapsychology databases.

4. The binary quality ratings used in parapsychological meta-analyses may also lead to insensitive quality analyses because they are crude measures of quality, whereas the seriousness of a flaw may often vary more smoothly than this in magnitude. For example, the use of card shuffling to randomize the target sequence in an ESP study would count as a flaw in most parapsychological meta-analyses; but, because randomness

improves as the number of shuffles increases, a study in which the deck was shuffled a lot would be less prone to error than a study in which the deck was shuffled only a few times. Analyses based on the usual binary flaw ratings may be too insensitive to pick up a relationship between flaws and effect size (Stanford & Stein, 1994).

5. Experimenters who obtain null results in their studies may give shorter accounts of them, leaving out details of the safeguards that they included (as Pratt, 1966, states that he did, for example). In a meta-analysis, such studies as a group would show a spurious association between low effect size and low quality; thus hiding, perhaps, a real association between low effect size and high quality in the other studies in the database (Milton, 1997).

6. Quality coding has almost always been conducted non-blind in parapsychology meta-analyses; so, it is difficult to rule out the possibility of coders being influenced in their coding by the studies' outcome. Coders who favor the psi hypothesis might be reluctant to ascribe flaws to successful studies or, conversely, might overcompensate for their bias by being more ready to penalize unsuccessful studies. Either strategy would introduce error variance.

7. Flaws might not behave additively but might instead interact with each other, reducing the sensitivity of simple contrast or correlation analyses that examine the relationship between total flaws and effect size (Stanford & Stein, 1994). Similarly, the relationship between the lack of any given safeguard and effect size might not be linear; a flaw may only become "active" above a certain threshold, for example, and, again, a simple correlative approach would be insensitive to this (Stanford & Stein, 1994).

8. If the presence of some flaws is negatively correlated, they might raise effect sizes in the database, but their effects would be difficult to detect. A database in which either safeguard A or safeguard B is present in each study, but never both together, serves as an extreme example to illustrate the point. If the absence of each safeguard increases effect size to roughly the same degree, then a comparison of effect sizes of studies in which safeguard A is present with studies in which it is absent will show no difference; nor will such a comparison show any difference when applied to safeguard B (Hyman, 1985).

There are plenty of reasons, then, for being cautious about concluding that methodological flaws do not increase study outcomes because estimates of studies' overall methodological quality are not statistically significantly correlated with effect size. Moreover, if the effects of flaws cannot be ruled out, then the other aspects of the meta-analyses' results that appear to support the psi hypothesis—that is, implausibly large "file drawers" of unpublished, null studies required to render the overall cumulation nonsignificant, and replicability across experimenters—also

are undermined. If study outcomes in a meta-analysis have been inflated by flaws, then so has the size of the "file drawer." If it were possible to correct study outcomes for the influence of those flaws, the overall cumulation would fall and the file drawer would, perhaps, no longer appear unreasonably large. Concerning replicability, all of the parapsychological databases that have examined it have shown statistically significant heterogeneity of effect size across studies (Honorton & Ferrari, 1989; Honorton, Ferrari & Bem, 1998; Milton, 1997; Radin & Ferrari, 1991; Radin & Nelson, 1989; Stanford & Stein, 1994), with the exception of the PRL ganzfeld database (Honorton et al., 1990). The replicability that many of them claim is replicability of successful rejection of the null hypothesis, using a variety of methods. Honorton and Ferrari (1989), for example, report that 30% of studies and 37% of experimenters obtained statistically significant results, indicating that more successful outcomes were obtained than the 5% expected by chance, and that success was not restricted to a few experimenters. Clearly, replicability defined in these terms is also vulnerable to explanation in terms of methodological artifacts in databases in which quality is unclear.

There is, however, a second type of evidence for psi that is often mentioned in addition to the results of proof-oriented meta-analyses, and that is that a number of literature reviews and meta-analyses of process-oriented psi research appear to indicate consistent relationships between study outcomes and variables such as belief in psi, extraversion, and so on. It appears to be often assumed that such relationships would not be consistent if they were attributable to methodological flaws. For example, it may be assumed that the "sheep-goat" effect, in which believers in psi score higher on psi tasks than nonbelievers, cannot be due to sensory cues because these cues would be equally available to both sheep and goats, and both groups would be expected to show the same level of performance.

This is not a safe assumption, however. There are many situations in which one might expect the action of flaws to produce consistent differences between groups, in line with parapsychologists' hypotheses. For example, in sheep-goat studies that do not have adequate sensory shielding, participants might be expected to be motivated to exploit those sensory cues (consciously or otherwise) to perform in accordance with their beliefs, just as they are hypothesized to do with extrasensory cues—sheep to score more hits and goats to score fewer hits. The pattern of the results due to the inadequate sensory shielding would mimic that expected under the usual sheep-goat hypothesis. As Palmer (1978) notes, the results of ESP experiments tend to fall into patterns that make psychological sense, inasmuch as they appear similar to the patterns of results that one might expect if subjects were attempting to respond to very weak sensory

information. Many spurious results due to flaws would also be expected to make similar sense, however, especially if in fact they were based on sensory leakage (see also Wiseman & Morris, 1995). Many of the more consistent findings of ESP research (such as higher scoring on confidence calls than on other trials [Carpenter, 1977; Palmer, 1978], higher scoring in studies with trial-by-trial feedback [Honorton & Ferrari, 1989], and so on) make conventional psychological sense if one assumes that they are due to the exploitation by participants of weaknesses in the design. Psychologically meaningful and consistent patterns of results would also be expected if safeguards preventing experimenter bias (such as predetermination of study sizes, prespecification of statistical tests, data checking and so on) were lacking. Arguing that process-oriented research has shown a consistent and meaningful pattern of results does not, therefore, allow side-stepping of the question of methodological quality if this argument is to be used in a proof-oriented way. Furthermore, it is difficult to make a strong proof-oriented case on the basis of this process-oriented work because meta-analyses of studies examining relationships between apparent ESP performance and moderator variables indicate similar problems of low or unclear quality in studies as are found in the proof-oriented meta-analyses (Honorton, Ferrari & Bem, 1998; Lawrence, 1993; Stanford & Stein, 1994).

I am not arguing that methodological problems clearly account for the positive results of the parapsychological meta-analyses. The study quality estimates that the meta-analyses report is in most cases minimum estimates of quality because they conservatively do not give the benefit of the doubt to studies that do not report details of safeguards; the actual quality of the studies may have been higher than it appears. The general absence of demonstrable relationships between studies' quality estimates and their effect sizes is encouraging for the psi hypothesis, if not a matter for complacency. Nor is it my intention to discourage the use of meta-analysis as a valuable tool because it cannot answer all of the questions that we would want to ask about a database. It is clearly a more powerful method than traditional literature reviews for synthesizing research findings; however, there appear to be potentially serious problems with drawing strong conclusions from reviews and meta-analyses of studies that are not demonstrably strong in quality, and these problems apply as much to process-oriented research as they do to proof-oriented research. If providing strong evidence for psi is still seen as important, then it appears that the only way to do so is by demonstrating a replicable, nonzero effect across a range of experimenters under stringent methodological conditions. So far, this does not appear to have happened.

Implications for future research

Ganzfeld research seems an obvious area in which to continue to look for strong evidence for psi. No other research methodology in parapsychology has received the detailed critical attention that ganzfeld research has received; it is the only area in which a whole database of studies has been examined intensively by a researcher such as Hyman who considers the existence of a genuine anomaly unlikely (Hyman, 1985) and in which researchers with opposing viewpoints have jointly produced methodological guidelines for research to settle the question of the existence of psi (Hyman & Honorton, 1986). In addition, it has arguably come to represent the case for psi in microcosm for mainstream psychology (Bem & Honorton, 1994; Milton & Wiseman, 1999a), and an account of it appears in every major summary of parapsychology's best evidence for psi (e.g., Atkinson, Atkinson, Smith, & Bem, 1990; Broughton, 1992, Hayes, 1998; Krippner et al., 1993; Radin, 1997; Utts, 1991). The failure of the recent studies to replicate the success of the earlier work therefore presents a challenge in the same mainstream scientific forum to parapsychology's claims for a genuine, replicable effect.

If ganzfeld research is to be an important player in the continued search for strong evidence, that search will only be successful if a replicable effect can be demonstrated. At present, however, if there is no change in the way ganzfeld research is carried out and no change in how replicability is examined, there appears to be no obvious reason why the next, inevitable meta-analysis of future ganzfeld studies will not show the same pattern of a null, or near-null, cumulation with perhaps a few individual experimenters obtaining effects that others are not replicating. In order to avoid repeating recent history, we need to know why the recent meta-analysis (Milton & Wiseman, 1999a) failed to replicate the findings of the PRL studies, which were carried out under similarly stringent conditions.

Unfortunately, the explanation is far from clear. One possible reason could be that the results of earlier ganzfeld studies were due to methodological problems rather than to psi; however, although a number of potential avenues for sensory leakage have been identified in the PRL work (Honorton et al., 1990; Morris, Cunningham, McAlpine, & Taylor, 1993; Wiseman, Smith, & Kornbrot, 1996), none appear sufficiently strong to account in any obvious way for the success of those studies, which were much more well-controlled than the earlier work (Milton & Wiseman, 1999a).

Another possibility is that the PRL studies used psi-conducive procedures but that the recent studies did not. This is possible but far from certain, for two reasons. First, although Bem and Honorton (1994) identified a number of variables that may be important for replication, the vast

majority of recent studies meta-analyzed (Milton & Wiseman, 1999a) either did not measure or did not report the average values of these variables in their studies (with the exception of the use of static versus dynamic targets where it is clear that the two databases are closely matched); and so it is not possible to make a strong case that differences in these variables accounted for the lack of replication (Milton & Wiseman, 1999a).

It is also possible that any number of additional, unidentified variables might have contributed to the success of the PRL studies; and so it is not possible to know whether the recent studies' failure to replicate the PRL work was due to their failure to exploit these variables to the same extent. There were a number of procedures used on all or almost all trials at PRL—the use of a sender, continuous auditory monitoring of the receiver's mentation by the sender, correspondence judging by the receiver rather than by an independent judge, (double-blind) prompting by the experimenter during the judging to correspondences that the receiver overlooked, a 14-minute pretrial relaxation procedure for both sender and receiver, and so on. The importance of these procedures has not been empirically determined. Any one or more of these procedures might be important for replication; however, without any evidence for their effects, it is not clear that the failure of the recent studies to replicate the findings of the PRL studies was due to the use of different procedures. It is not evident, at this point, what a replication of the PRL work in its essentials would have to consist of.

Since the convention presentation of our meta-analysis (Milton & Wiseman, 1997a), a number of colleagues have informally suggested that if we had restricted our database to "standard" ganzfeld studies (i.e., studies without unusual features) across-experimenter replication of the PRL effect size might have been evident. However, among the researchers who have discussed the issue with me there appears to be little agreement about the features of a standard ganzfeld study. Devising a rule to define such a study at this point could easily appear as a post hoc attempt to explain away a disappointing result, given that the previous ganzfeld meta-analyses included almost all studies and trials no matter how unusual their procedures (Bem & Honorton, 1994; Honorton, 1985; Hyman, 1985) and regardless of whether those procedures would be expected to result in success or failure.[2] Neither Hyman and Honorton (1986) nor Bem and Honorton (1994) specified that studies would have to have certain features to be considered part of the replicability test that they proposed. It does not appear possible to selectively meta-analyze the recent studies and make a strong case that the ganzfeld effect is replicable; however, a selective meta-analysis with exclusion criteria stated in advance of studies being conducted would be a credible demonstration of replicability if it obtained positive results. In practice, it is

unlikely that criteria could be set up that would anticipate all of the novel features that experimenters might introduce in their studies that would lead most researchers to expect them to be unsuccessful. In addition to having to conform to a basic set of criteria, the procedures planned for each study would therefore also have to be examined on a case-by-case basis to determine whether or not the study ought to be included in the replication test. The existence of such a project would neither affect the usual conduct of process-oriented research nor force experimenters to use certain procedures in their studies. It would simply be the case that studies eligible to be included in the meta-analysis would be included and others would not. Similarly, the project would not affect anyone's usual freedom to conduct a meta-analysis of their own. In particular, there is no reason anyone should not conduct a process-oriented meta-analysis involving all studies.

Some researchers may believe that it is already possible to identify successful ganzfeld studies based on their procedures alone, and that it would be advisable to begin such a meta-analysis now. Others may think this premature. Very few variables have been explored repeatedly or systematically in ganzfeld studies, and even fewer have been examined meta-analytically across studies to determine whether there is good statistical evidence that they relate to effect size. Meta-analytic investigation of some of the variables suggested by Bem and Honorton (1994) as having been important in the PRL work indicates that other experimenters have not replicated their effects in the few areas where this has been attempted (Milton & Wiseman, 1999a). In addition, some variables identified by Bem and Honorton as having had statistically significant relationships with effect size in the PRL studies do not in fact appear to have done so (Milton & Wiseman, 1999a), suggesting that our success so far in identifying what variables are important in the ganzfeld might be more limited

---

[2] The previous ganzfeld meta-analyses did not report explicit exclusion rules but the implicit rules appear to have been to include every ganzfeld study (for Hyman's meta-analysis) or every single trial (for the PRL meta-analysis) in which a ganzfeld environment (even a modified one) was used to conduct an ESP test, with one disputed exception. For the first meta-analysis of ganzfeld studies, Honorton provided Hyman with "a copy of every ganzfeld study known to him" (Hyman, 1985, p. 4), all of which Hyman included in his meta-analysis. The studies were procedurally very varied, with some having features that laboratory lore might predict would not be psi-conducive, such as very short mentation periods (e.g. Rogo et al., 1976); however, Honorton did exclude two conditions in a study by Raburn (1975) in which participants were not aware that they were taking part in an ESP test, on the grounds that these trials were too atypical of other ganzfeld research. Hyman (1985) objected to their exclusion because other studies contained unique features and yet were included in the database. Bem and Honorton's (1994) subsequent meta-analysis of the PRL work included every single trial done using the autoganzfeld. The PRL studies were also procedurally varied and the meta-analysis included trials that, again, might arguably not be expected to be successful, such as demonstration trials carried out in the presence of a TV crew and trials from Series 302 in which Target 79 was included in the target set on each trial despite its never having been previously correctly identified when serving as the target.

than has been assumed. Before embarking upon a replication test that should exploit its findings, it may be that a systematic assessment of process-oriented ganzfeld research is called for (e.g., see Dalton, 1997b).

Summary and conclusion

The meta-analysis of recent, well-controlled ganzfeld studies (Milton & Wiseman, 1999a) indicates a failure to replicate the results of the earlier work, and the evidence for psi from meta-analyses and process-oriented reviews of parapsychology studies of low or uncertain quality does not appear compelling. If the search for strong evidence for psi is to continue, ganzfeld research appears to be its natural arena. A meta-analysis that excludes studies before they are conducted if they are not expected to replicate a positive effect appears to be the obvious test of future replication. Until more research has been done to identify what factors may be psi conducive in the ganzfeld, such a meta-analysis may be premature, but it appears to be an important goal to work towards.

Many researchers may disagree with my assessment of the evidence for psi accumulated so far, and with my goal of continuing to seek stronger evidence in general, and with my proposal for a prospective ganzfeld meta-analysis in particular. Conversely, many may disagree with the use of meta-analyses of studies of uncertain quality being promoted as strong evidence for psi, and with ganzfeld research having become a crucial test case before the factors that affect its replicability have been well-established. Whatever researchers' views may be, however, the momentum of previous events is carrying the field towards another inclusive meta-analysis of future ganzfeld studies that appears likely to show the same failure to replicate as did the last one. Should a second failure to replicate occur despite the warning of a first failure, it will give the appearance of reasonably strong evidence against claims for psi as a replicable (and therefore, probably genuine) effect.

If this is not a direction that parapsychologists want events to take, then now appears to be the time to say so. Although the choice of whether to carry out a meta-analysis is likely to be an individual one, its results will affect other researchers. The opportunity for the research community, rather than a few, key individuals, to discuss the issues and express their opinions is long overdue. I look forward to hearing the views of my colleagues on the matters that I have discussed in this paper.

Organization of an electronic mail discussion

The apparent replication problems in ganzfeld research described in the preceding paper appeared to require discussion among the ganzfeld research community in order to determine what, if any, course of action seemed appropriate and could be agreed upon. I, therefore, invited a group of researchers with expertise in ganzfeld research and

parapsychological meta-analysis to discuss the future of ganzfeld research in a three-week electronic mail conference in May 1999.

The invitees consisted of authors and coauthors of studies conducted since publication of the Hyman-Honorton guidelines; Honorton's autoganzfeld research team; senior authors of at least two ganzfeld studies conducted at any time; meta-analysts of ganzfeld research and of other proof-oriented meta-analyses; authors of published commentary on proof-oriented aspects of the ganzfeld meta-analyses; and, in order to include future ganzfeld experimenters, researchers planning to conduct a formal ganzfeld study within the next two years.

Every effort was made to identify and locate eligible participants. Researchers planning to conduct ganzfeld studies within two years were sought via messages on the two main parapsychology electronic mailbases (PRF and PDL). For participants eligible through previous authorship, contact details were sought from these mailbases, the Parapsychological Association, alumni offices in UK universities (for those who had conducted research while students), former colleagues and co-authors, internet directory searches, and other invitees. Out of 65 eligible participants, 58 were successfully traced. Forty-one invitees (71%) accepted the invitation to join the mailbase during its operation. They are listed in Appendix B. Each received an advance copy of the preceding discussion paper and a preview copy of the Milton and Wiseman (1999a) ganzfeld meta-analysis paper, then in press with *Psychological Bulletin*.

Because of the importance of the issues under discussion, John Palmer, the editor of this journal, agreed in advance to publish a transcript of the debate. Participants were informed of this at the time of their invitation. They were also told that the transcript would be edited for length and re-ordered if necessary by an independent editor, Dr. Gertrude Schmeidler. They were informed that any editing would be agreed with each message's author before publication and that, to avoid bias, no substantive content would be removed. Participants were assured that in the interests of neutrality I would have no involvement in this editing and that John Palmer, himself a debate participant, would restrict himself to approving the edited material's length and would have no influence on the nature of its content.

The debate format had some unusual features intended to foster productive discussion. Participants were informed that there would be a strict policy of courtesy among discussants. In addition, so that arguments would be assessed on their merit rather than on their author's status, each author's identification and e-mail address were removed by a computer program en route to the mailbase and each message was only identified by a number. Authors could, however, reveal their identities in any particular message if this was necessary to make it clear that they

spoke with authority on a question of fact (for example, in discussing unpublished data from their own research). Otherwise, participants were asked to help conceal their identities by wording messages in ways that would not reveal who they were. Participants were informed that the identity of each message's author would be announced after the discussion had closed and would be published with the debate transcript.

In order to ensure compliance with the rules of the discussion, a moderator, Professor Hoyt Edge, screened each message for anonymity and courtesy, with a remit to negotiate if necessary an acceptable wording before posting the message on to the other participants. Participants were informed that I would have no involvement in the moderating process, again in the interests of neutrality.

All members of the discussion group received an optional questionnaire before and after the discussion asking their opinions on the main issues, and a post-discussion questionnaire concerning their satisfaction with the organizational features of the debate. The questionnaire data are presented in Appendix C.

The edited debate material follows in Part II. Each message in the transcript is numbered, with its author listed in an appendix so that readers may, if they wish, have the same experience as the discussants of reading the material without knowing who wrote it, allowing themselves only to be swayed by the force of argument and evidence.

References

Atkinson, R. L., Atkinson, R. C., Smith, E. E., & Bem, D. J. (1990). *Introduction to psychology* (10th ed.). Orlando, FL: Harcourt Brace Jovanovich.

Bem, D. J., & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin*, **115**, 4-18.

Broughton, R. (1992). *Parapsychology: The controversial science.* London: Rider.

Carpenter, J. C. (1977). Intrasubject and subject-agent effects in ESP experiments. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 202-272). New York: Van Nostrand Reinhold.

Dalton, K. (1997a). Exploring the links: Creativity and psi in the ganzfeld. *Proceedings of Presented Papers: The Parapsychological Association 40th Annual Convention*, 119-134.

Dalton, K. (1997b). Is there a formula to success in the ganzfeld? Observations on predictors of psi-ganzfeld performance. *European Journal of Parapsychology*, **13**, 71-82.

Hayes, N. (1998). *Foundations of psychology: An introductory text* (2nd ed.). Walton-on-Thames, England: Nelson.

Honorton, C. (1985). Meta-analysis of psi ganzfeld research: A response to Hyman. *Journal of Parapsychology*, **49**, 51-91.

Honorton, C., Berger, R. E., Varvoglis, M. P., Quant, M., Derr, P., Schechter, E. I., & Ferrari, D. C. (1990). Psi communication in the

ganzfeld: Experiments with an automated testing system and a comparison with a meta-analysis of earlier studies. *Journal of Parapsychology*, **54**, 99-139.

HONORTON, C., & FERRARI, D. C. (1989). Meta-analysis of forced-choice precognition experiments. *Journal of Parapsychology*, **53**, 281-308.

HONORTON, C., FERRARI, D.C. & BEM, D.J. (1998). Extraversion and ESP performance: A meta-analysis and a new confirmation. *Journal of Parapsychology*, **62**, 255-276.

HYMAN, R. (1985). The ganzfeld psi experiment: A critical appraisal. *Journal of Parapsychology*, **49**, 3-49.

HYMAN, R., & HONORTON, C. (1986). A joint communiqué: The psi ganzfeld controversy. *Journal of Parapsychology*, **50**, 350-364.

KRIPPNER, S., BRAUD, W., CHILD, I. L., PALMER, J., RAO, K. R., SCHLITZ, M., WHITE, R. A., & UTTS, J. (1993). Demonstration research and meta-analysis in parapsychology. *Journal of Parapsychology*, **57**, 275-286.

LAWRENCE, T. (1993). Gathering in the sheep and goats . . . A meta-analysis of forced-choice sheep-goat ESP studies, 1947-1993. *Proceedings of Presented Papers: The Parapsychological Association 36th Annual Convention*, 75-86.

MILTON, J. (1997). Meta-analysis of free-response studies without altered states of consciousness. *Journal of Parapsychology*, **61**, 279-319.

MILTON, J., & WISEMAN, R. (1997a). Ganzfeld at the crossroads: A meta-analysis of the new generation of studies. *Proceedings of Presented Papers: The Parapsychological Association 40th Annual Convention*, 267-282.

MILTON, J., & WISEMAN, R. (1997b). *Guidelines for extrasensory perception research.* Hatfield, England: University of Hertfordshire Press.

MILTON, J., & WISEMAN, R. (1999a). Does psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin*, **125**, 387-391.

MILTON, J., & WISEMAN, R. (1999b). A meta-analysis of mass-media tests of extrasensory perception. *British Journal of Psychology*, **90**, 235-240.

MORRIS, R. L., CUNNINGHAM, S., MCALPINE, S., & TAYLOR, R. (1993). Toward replication and extension of autoganzfeld results. *Proceedings of Presented Papers: The Parapsychological Association 36th Annual Convention*, 177-191.

PALMER, J. (1978). Extrasensory perception: Research findings. In S. Krippner (Ed.), *Advances in parapsychological research 2: Extrasensory perception*, (pp. 59-243). New York: Plenum Press.

PARKER, A., & WESTERLUND, J. (1998). Current research in giving the ganzfeld an old and a new twist. *Proceedings of Presented Papers: The Parapsychological Association 41st Annual Convention*, 135-142.

PRATT, J. G. (1966). New ESP tests with Mrs. Gloria Stewart. *Journal of the American Society for Psychical Research*, **60**, 321-339.

RABURN, L. (1975). *Expectation and transmission factors in psychic functioning.* Unpublished honors thesis, Tulane University, New Orleans, LA.

RADIN, D. I. (1997). *The conscious universe: The scientific truth of psychic phenomena.* New York, NY: HarperCollins.

RADIN, D. I., & FERRARI, D. C. (1991). Effects of consciousness on the fall of dice: A meta-analysis. *Journal of Scientific Exploration*, **5**, 61-83.

RADIN, D. I., & NELSON, R. D. (1989). Evidence for consciousness-related anomalies in random physical systems. *Foundations of Physics*, **19**, 1499-1514.

ROCO, D. S., SMITH, M., & TERRY, J. (1976). The use of short-duration ganzfeld stimulation to facilitate psi-mediated imagery. *European Journal of Parapsychology*, **1 (3)**, 72-77.

ROSENTHAL, R. (1978). How often are our numbers wrong? *American Psychologist*, **33**, 1005-1008.

STANFORD, R. G., & STEIN, A. G. (1994). A meta-analysis of ESP studies contrasting hypnosis and a comparison condition. *Journal of Parapsychology*, **58**, 235-269.

STEINKAMP, F., MILTON, J., & MORRIS, R. L. (1998). A meta-analysis of forced-choice experiments comparing clairvoyance and precognition. *Journal of Parapsychology*, **62**, 193-218.

SYMMONS, C., & MORRIS, R. L. (1997). Drumming at seven Hz and automated ganzfeld performance. *Proceedings of Presented Papers: The Parapsychological Association 40th Annual Convention*, 441-453.

UTTS, J. (1991). Replication and meta-analysis in parapsychology. *Statistical Science*, **6**, 363-403.

WEZELMAN, R., & BIERMAN, D. J. (1997). Process oriented ganzfeld research in Amsterdam Series IV B (1995): emotionality of target material, Series V (1996) and Series VI (1997): judging procedure and altered states of consciousness. *Proceedings of Presented Papers: The Parapsychological Association 40th Annual Convention*, 441-453.

WEZELMAN, R., GERDING, J. L. F., & VERHOEVEN, I. (1997). Eigensender ganzfeld psi: An experiment in practical philosophy. *European Journal of Parapsychology*, **13**, 28-39.

WISEMAN, R., & MORRIS, R. L. (1995). *Guidelines for testing psychic claimants.* Amherst, NY: Prometheus.

WISEMAN, R., SMITH, D., & KORNBROT, D. (1996). Exploring possible sender-to-experimenter acoustic leakage in the PRL autoganzfeld experiments. *Journal of Parapsychology*, **60**, 97-128.

Department of Psychology
University of Edinburgh
7 George Square
Edinburgh EH8 9JZ
Scotland, UK

APPENDIX A

TABLE A1

GANZFELD STUDIES PUBLISHED TO DATE (MARCH 1999) SINCE COMPLETION OF MILTON & WISEMAN (1999A) META-ANALYSIS (FEBRUARY 1997)

Study (N = 12)	Number of trials	*z*	$z/N^{1/2}$
Dalton (1997a)	128	5.26	.46
Parker & Westerlund (1998) Study IV	30	2.40	.44
Parker & Westerlund (1998) Study V	30	1.25	.23
Parker & Westerlund (1998) Serial Ganzfeld	30	—[a]	—[a]
Symmons & Morris (1997) Pilot Study	12	—[b,c]	—[b,c]
Symmons & Morris (1997) Main Study	51	2.98[b]	.42[b]
Wezelman & Bierman (1997) Amsterdam Series IV B	32	-1.48	-.26
Wezelman & Bierman (1997) Amsterdam Series V	40	-.91	-.14
Wezelman & Bierman (1997) Amsterdam Series VI	40	-.15	-.2
Wezelman & Bierman (1997) Amsterdam Series VI Exploratory Meditation Trials	7	-1.11	-.42
Wezelman & Bierman (1997) Amsterdam Series VI Exploratory Psilocybine Trials	12	—[d]	—[d]
Wezerman et al. (1997)	32	2.15	.38

[a] In this study, the receiver's task was to place the four targets in the judging set in the order in which they had been presented during the ganzfeld session. The authors present the results as a frequency table of the number of correct placements within each trial. By inspection the outcome is slightly below chance; however, the authors do not present or refer to any specific inferential statistical analysis and, because it is not clear what analysis was intended, no post hoc analysis has been imposed here.

[b] In both studies by Symmons and Morris, tapes of drumming at different frequencies were used instead of white noise, and so it is questionable whether they can be considered as using a ganzfeld environment. The studies are included here to make clear the effects on the database of including or excluding them.

[c] No outcome was reported for the pilot trials.

[d] Two receivers guessed at the same target on 6 trials, obtaining 7 hits in the resulting 12 trials; however, data are not presented that would allow for correction of the nonindependence of their calls (the "stacking effect": see Milton & Wiseman, 1999b), and so no outcome is presented here.

[e] This table contains an erratum. The effect size for Wezelman & Bierman (1997) Amsterdam Series VI should be -.02.

Table A2

POST HOC COMPARISONS BETWEEN MEAN EFFECT SIZES IN META-ANALYSES OF RECENT AND EARLIER GANZFELD STUDIES

Databases compared	*t*	d.f.	*p*(one-tailed)
Honorton (1985) vs.:			
Milton & Wiseman (1999a)	3.06	56	.0017
All studies 1987 to present	2.90	65	.0026
All studies 1987 to present excl. Dalton (1997a)	3.04	64	.0017
Bem & Honorton (1994) vs.:			
Milton & Wiseman (1999a)	2.64	39	.0059
All studies 1987 to present	2.22	48	.016
All studies 1987 to present excl. Dalton (1997a)	2.38	47	.011

APPENDIX B

Members of the Discussion Group

Members of the discussion group, in alphabetical order, were as follows (those who posted messages are marked with an asterisk): Cheryl Alexander, Daryl Bem, Dick Bierman,* Douwe Bosga, William Braud,* Kathy Dalton, Deborah Delanoy, Norman Don, Ricardo Eppinger, Hans Gerding,* Gerd Hövelmann,* Anjum Khilji, Diana Kornbrot, Tony Lawrence,* Bruce McDonough,* Stuart Menzies,* Julie Milton,* Bob Morris,* Roger Nelson,* John Palmer,* Adrian Parker,* Dean Radin,* Chris Roe, Ephraim Schechter,* Marilyn Schlitz, Fabio da Silva, Matthew Smith, Rex Stanford, Fiona Steinkamp,* Charles Symmons,* James Terry, Jessica Utts, Mario Varvoglis, Charles Warren, Caroline Watt,* Joakim Westerlund, Rens Wezelman,* Nils Wiklund,* Carl Williams, Melvyn Willin,* Richard Wiseman.

APPENDIX C

Questionnaire Data

As noted earlier, all members of the mailbase group were sent an optional pre- and postdiscussion questionnaire concerning the main issues, and a postdiscussion questionnaire asking about their satisfaction with the organizational features of the debate. To minimize response bias, discussants were asked to send their responses for compilation to the moderator, who would keep their individual replies permanently confidential from me.

Pre- and Postdiscussion Opinions on the Main Issues

The results of the questionnaires are summarized in Table 1. Just under half of the mailbase members answered the pre- and postdiscussion questionnaires, and so it is not clear that the results proportionately reflect the views of whole group. Respondents were not asked to give their identities, to maximize response rates. It is, therefore, not clear whether any change in opinion reflects a change in the opinion of broadly the same group of people, or a change in the identities of those responding to the questionnaire. The data can only be interpreted as reflecting the views of those who chose to express an opinion at the time.

Bearing these limitations in mind, it can be seen that respondents appeared to maintain their position of tending to favor (but with some uncertainty) the view that the experimental evidence for psi as a genuine anomaly is strong enough to convince a neutral scientist. Respondents tended to agree before the discussion that ganzfeld research should continue as an important focus for psi as a genuine effect, replicable across experimenters under certain conditions; and they agreed more strongly with this view after the discussion. There was little change in respondents' view that meta-analyses of stringently conducted studies are important as part of the case for psi as a replicable, genuine anomaly, nor in their view that it is necessary to plan exclusions in advance rather than post hoc in the next ganzfeld meta-analysis. Before the debate, respondents had a slight tendency to believe on balance that it is already possible to identify successful ganzfeld studies reasonably reliably in advance on the basis of their procedures; but afterwards the majority did not think this possible.

Table 3
OPINIONS ON THE MAIN DISCUSSION ISSUES BEFORE AND AFTER DEBATE

		Percent Agreement	
Question	Response	Entry Poll (N = 16)[a]	Exit Poll (N = 18)[b]
1. Do you think that the experimental evidence for psi is strong enough that a neutral scientist should be convinced that a genuine anomaly has been demonstrated, that is, that there is a phenomenon not explicable in terms of error, selective reporting, fraud, ordinary sensorimotor effects and so on?	Yes, certainly	13	6
	Yes, on balance	*50*	*44*
	Uncertain	*31*	*39*
	No, on balance	*0*	*0*
	No, certainly not	6	11
2. Do you think that ganzfeld research should remain an important focus for testing the hypothesis that, at least under certain conditions, psi is a genuinely anomalous effect that can be replicated across experimenters?	I do not believe that further testing of this hypothesis is necessary, it has already been sufficiently confirmed	13	6
	No, certainly not	*13*	*0*
	No, on balance	*13*	*11*
	Uncertain	*19*	*11*
	Yes, on balance	*19*	*61*
	Yes, certainly	25	11
3a. How important do you think meta-analyses of stringently conducted parapsychology studies are in making at least part of the case for psi as a genuine and replicable anomaly?	Crucial	13	12
	Important	*63*	*59*
	Uncertain	*13*	*6*
	Not important	*6*	*6*
	Irrelevant	6	18

Discussion Paper 333

3b[c]. I am proposing that if a meta-analysis of ganzfeld studies designed to test whether psi is a genuine and replicable anomaly is to be selective and yet still credible, it is necessary to identify studies for inclusion in advance of their conduct on the basis of their planned procedures rather than excluding studies after they have been conducted when their results are known. Do you agree?	Yes, certainly	29	40
	Yes, on balance	*36*	*33*
	Uncertain	*7*	*13*
	No, on balance	*21*	*13.*
	No, certainly not	7	0
4. Do you think that the procedures necessary for producing a replicable ganzfeld effect have been identified to the extent that it would be possible now to identify in advance which studies are likely to be successful with reasonable reliability?	No, certainly not	13	28
	No, on balance	*25*	*28*
	Uncertain	*19*	*28*
	Yes, on balance	*38*	*17*
	Yes, certainly	6	0

[a] N = 14 for Question 3b

[b] N = 17 for Question 3a and N = 15 for Question 3b

[c] This question was only for respondents who answered "crucial" or "important" to Question 3a.

Opinions on Discussion Features

Seventeen members of the discussion group responded to the postdiscussion questionnaire asking for their views on various aspects of how the discussion was run. Concerning the time allowed for the discussion, most (65%) thought three weeks to be about right. The remainder (35%) would have preferred a longer period (between four and eight weeks, according to individual responses), with none thinking the debate too long. Most respondents would have recommended message anonymity and prearranged publication for future e-mail debates (70% and 69%, respectively), with a moderator to screen for courtesy being very strongly favored: 83% of respondents would have recommended this feature for a future debate. The present discussion's moderator was not given the role of guiding the discussion, but 59% of respondents recommended such guidance for future e-mail debates.

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

Part IV
Psychokinesis and Distant Mental Influence

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[18]

EXPERIMENTAL INVESTIGATION OF TABLE-MOVING

M. Faraday

PROFESSOR FARADAY ON TABLE-MOVING.

THE following account of the methods pursued and the results obtained by Prof. Faraday in the investigation of a subject which has taken such strange occupation of the public mind, both here and abroad, has been communicated to our columns by that high scientific authority. The subject was generally opened by Mr. Faraday in the *Times* of Thursday: it being therein intimated that the details were to be reserved for our this day's publication. The communication is of great importance in the present morbid condition of public thought,—when, as Professor Faraday says, the effect produced by table-turners has, without due inquiry, been referred to electricity, to magnetism, to attraction, to some unknown or hitherto unrecognized physical power able to affect inanimate bodies, to the revolution of the earth, and even to diabolical or supernatural agency:—and we are tempted to extract a passage from Mr. Faraday's letter to the *Times* which we think well worth adding to the experimental particulars and the commentaries with which he has favoured ourselves. "I have been," says the Professor, "greatly startled by the revelation which this purely physical subject has made of the condition of the public mind. No doubt, there are many persons who have formed a right judgment or used a cautious reserve,—for I know several such, and public communications have shown it to be so; but their number is almost as nothing to the great body who have believed and borne testimony, as I think, in the cause of error. I do not here refer to the distinction of those who agree with me and those who differ. By the great body, I mean such as reject all consideration of the equality of cause and effect,—who refer the results to electricity and magnetism, yet know nothing of the laws of these forces,—or to attraction, yet show no phenomena of pure attractive power,—or to the rotation of the earth, as if the earth revolved round the leg of a table,—or to some unrecognized physical force, without inquiring whether the known forces are not sufficient,—or who even refer them to diabolical or supernatural agency, rather than suspend their judgment, or acknowledge to themselves that they are not learned enough in these matters to decide on the nature of the action. I think the system of education that could leave the mental condition of the public body in the state in which this subject has found it must have been greatly deficient in some very important principle."

Experimental Investigation of Table-Moving.

The object which I had in view in this inquiry was not to satisfy myself, for my conclusion had been formed already on the evidence of those who had turned tables,—but that I might be enabled to give a strong opinion, founded on facts, to the many who applied to me for it. Yet, the proof which I sought for, and the method followed in the inquiry, were precisely of the same nature as those which I should adopt in any other physical investigation. The parties with whom I have worked were very honourable, very clear in their intentions, successful table-movers, very desirous of succeeding in establishing the existence of a peculiar power, thoroughly candid, and very effectual. It is with me a clear point that the table moves when the parties, though they strongly wish it, do not intend, and do not believe that they move it by ordinary mechanical power. They say, the table draws their hands; that it moves first, and they have to follow it,—that sometimes it even moves from under their hands. With some the table will move to the right or left according as they wish or will it,—with others the direction of the first motion is uncertain:—but all agree that the table moves the hands and not the hands the table. Though I believe the parties do not intend to move the table, but obtain the result by a *quasi* involuntary action,—still I had no doubt of the influence of expectation upon their minds, and through that upon the success or failure of their efforts. The first point, therefore, was, to remove all objections due to expectation, having relation to the substances which I might desire to use:—so, plates of the most different bodies, electrically speaking, —namely, sand-paper, millboard, glue, glass, moist clay, tinfoil, cardboard, gutta percha, vulcanized rubber, wood, &c.,—were made into a bundle and placed on a table under the hands of a turner. The table turned. Other bundles of other plates were submitted to different persons at other times,—and the tables turned. Henceforth, therefore, these substances may be used in the construction of apparatus. Neither during their use nor at other times could the slightest trace of electrical or magnetic effects be obtained. At the same trials it was readily ascertained that one person could produce the effect; and that the motion was not necessarily circular, but might be in a straight line. No form of experiment or mode of observation that I could devise gave me the slightest indication of any peculiar natural force. No attractions, or repulsions, or signs of tangential power, appeared,—nor anything which could be referred to other than the mere mechanical pressure exerted inadvertently by the turner. I therefore proceeded to analyze this pressure, or that part of it exerted in a horizontal direction:—doing so, in the first instance, unawares to the party. A soft cement, consisting of wax and turpentine, or wax and pomatum, was prepared. Four or five pieces of smooth slippery cardboard were attached one over the other by little pellets of the cement, and the lower of these to a piece of sand-paper resting on the table; the edges of these sheets overlapped slightly, and on the under surface a pencil line was drawn over the laps so as to indicate position.

The upper cardboard was larger than the rest, so as to cover the whole from sight. Then, the table-turner placed the hands upon the upper card,—and we waited for the result. Now, the cement was strong enough to offer considerable resistance to mechanical motion, and also to retain the cards in any new position which they might acquire,—and yet weak enough to give way slowly to a continued force. When at last the tables, cards, and hands all moved to the left together, and so a true result was obtained, I took up the pack. On examination, it was easy to see by the displacement of the parts of the line, that the hand had moved further than the table, and that the latter had lagged behind;—that the hand, in fact, had pushed the upper card to the left, and that the under cards and the table had followed and been dragged by it. In other similar cases when the table had not moved, still the upper card was found to have moved, showing that the hand had carried it in the expected direction. It was evident, therefore, that the table had not drawn the hand and person round, nor had it moved simultaneously with the hand. The hand had left all things under it behind, and the table evidently tended continually to keep the hand back.

The next step was, to arrange an index, which should show whether the table moved first, or the hand moved before the table, or both moved or remained at rest together. At first this was done by placing an upright pin fixed on a leaden foot upon the table, and using that as the fulcrum of a light lever. The latter was made of a slip of foolscap paper, and the short arm, about ¼ of an inch in length, was attached to a pin proceeding from the edge of a slipping card placed on the table, and prepared to receive the hands of the table-turner. The other arm, of 11½ inches long, served for the index of motion. A coin laid on the table marked the normal position of the card and index. At first the slipping card was attached to the table by the soft cement, and the index was either screened from the turner, or the latter looked away: then, before the table moved, the index showed that the hand was giving a resultant pressure in the expected direction. The effect was never carried far enough to move the table, for the motion of the index corrected the judgment of the experimenter, who became aware that, inadvertently, a side force had been exerted. The card was now set free from the table, *i. e.*, the cement was removed. This, of course, could not interfere with any of the results expected by the table-turner,—for both the bundle of plates spoken of and single cards had been freely moved on the tables before; but now that the index was there, witnessing to the eye, and through it to the mind, of the table-turner, not the slightest tendency to motion either of the card or of the table occurred. Indeed, whether the card was left free or attached to the table all motion or tendency to motion was gone. In one particular case there was relative motion between the table and the hands: I believe that the hands moved in one direction; the table-turner was persuaded that the table moved from under the hand in the other direction:—a gauge, standing upon the floor, and pointing to the table, was therefore set up on that and some future occasions,—and then, neither motion of the hand nor of the table occurred.

A more perfect lever apparatus was then constructed in the following manner:—Two thin boards, 9½ inches by 7 inches, were provided; a board, 9 by 5 inches, was glued to the middle of the underside of one of these, (to be called the table-board), so as to raise the edges free from the table; being placed on the table, near and parallel to its side, an upright pin was fixed close to the further edge of the board, at the middle, to serve as the fulcrum for the indicating lever. Then, four glass rods, 7 inches long and ¼ in diameter, were placed as rollers on different parts of this table-board, and the upper board placed on them; the rods permitted any required amount of pressure on the boards, with a free motion of the upper on the lower to the right and left. At the part corresponding to the pin in the lower board, a piece was cut out of the upper board, and a pin attached there which, being bent downwards, entered the hole in the end of the short arm of the index lever: this part of the lever was of cardboard; the indicating prolongation was a straight hay-stalk 15 inches long. In order to restrain the motion of the upper board on the lower, two vulcanized rubber rings were passed round both, at the parts not resting on the table: these, whilst they tied the boards together, acted also as springs,—and whilst they allowed the first feeblest tendency to motion to be seen by the index, exerted before the upper board had moved a quarter of an inch sufficient power in pulling the upper board back from either side, to resist a strong lateral action of the hand. All being thus arranged, except that the lever was away,—the two boards were tied together with string, running parallel to the vulcanized rubber springs, so as to be immovable in relation to each other. They were then placed on the table, and a table-turner sat down to them:—the table very shortly moved in due order, showing that the apparatus offered no impediment to the action. A like apparatus, with metal rollers, produced the same result under the hands of another person. The index was now put into its place and the string loosened, so that the springs should come into play. It was soon seen, with the party that could will the motion in either direction, (from whom the index was purposely hidden,) that the hands were gradually creeping up in the direction before agreed upon, though the party certainly thought they were pressing downwards only. When shown that it was so, they were truly surprised; but when they lifted up their hands and immediately saw the index return to its normal position, they were convinced. When they looked at the index and could see for themselves whether they were pressing truly downwards, or obliquely so as to produce a resultant in the right or left handed direction, then such an effect never took place. Several tried, for a long while together, and with the best will in the world; but no motion, right or left, of the table, or hand, or anything else, occurred.—[A passage from the letter in the *Times* is worth reproducing here,—as illustrating in other words the value of this method of self-conviction.—"The result," says Prof. Faraday, "was, that when the parties saw the index it remained very steady; when it was hidden from them, or they looked away from it, it wavered about, though they believed that they always pressed directly down-

wards; and, when the table did not move, there was still a resultant of hand force in the direction in which it was wished the table should move, which, however, was exercised quite unwittingly by the party operating. This resultant it is which, in the course of the waiting time, while the fingers and hands become stiff, numb, and insensible by continued pressure, grows up to an amount sufficient to move the table or the substances pressed upon. But the most valuable effect of this test-apparatus (which was afterwards made more perfect and independent of the table) is the corrective power it possesses over the mind of the table-turner. As soon as the index is placed before the most earnest, and they perceive—as in my presence they have always done — that it tells truly whether they are pressing downwards only or obliquely, then all effects of table-turning cease, even though the parties persevere, earnestly desiring motion, till they become weary and worn out. No prompting or checking of the hands is needed—*the power is gone;* and this only because the parties are made conscious of what they are really doing mechanically, and so are unable unwittingly to deceive themselves. I know that some may say that it is the cardboard next the fingers which moves first, and that *it* both drags the table and also the table-turner with it. All I have to reply is, that the cardboard may in practice be reduced to a thin sheet of paper weighing only a few grains, or to a piece of goldbeaters' skin, or even the end of the lever, and (in principle) to the very cuticle of the fingers itself. Then the results that follow are too absurd to be admitted: the table becomes an incumbrance, and a person holding out the fingers in the air, either naked or tipped with goldbeaters' skin or cardboard, ought to be drawn about the room, &c.; but I refrain from considering imaginary yet consequent results which have nothing philosophical or real in them."]

Another form of index was applied thus:—a circular hole was cut in the middle of the upper board, and a piece of cartridge paper pasted under it on the lower surface of the board; a thin slice of cork was fixed on the upper surface of the lower board corresponding to the cartridge paper; the interval between them might be a quarter of an inch or less. A needle was fixed into the end of one of the index hay-stalks, and when all was in place the needle point was passed through the cartridge paper and pressed slightly into the cork beneath, so as to stand upright: then any motion of the hand, or hand-board, was instantly rendered evident by the deflection of the perpendicular hay-stalk to the right or left.

I think the apparatus I have described may be useful to many who really wish to know the truth of nature, and would prefer that truth to a mistaken conclusion: desired, perhaps, only because it seems to be new or strange. Persons do not know how difficult it is to press directly downward, or in any given direction against a fixed obstacle: or even to *know only* whether they are doing so or not; unless they have some indicator, which, by visible motion or otherwise, shall instruct them: and this is more especially the case when the muscles of the fingers and hand have been cramped and rendered either tingling, or insensible, or cold by long continued pressure. If a finger be pressed constantly into the corner of a window frame for ten minutes or more, and then, continuing the pressure, the mind be directed to judge whether the force at a given moment is all horizontal, or all downward, or how much is in one direction and how much in the other, it will find great difficulty in deciding; and will at last become altogether uncertain: at least such is my case. I know that a similar result occurs with others; for I have had two boards arranged, separated, not by rollers but by plugs of vulcanized rubber, and with the vertical index: when a person with his hands on the upper board is requested to press only downwards, and the index is hidden from his sight, it moves to the right, to the left, to him and from him, and in all horizontal directions; so utterly unable is he strictly to fulfil his intention without a visible and correcting indicator. Now, such is the use of the instrument with the horizontal index and rollers: the mind is instructed, and the involuntary or *quasi* involuntary motion is checked in the commencement, and therefore never rises up to the degree needful to move the table, or even permanently the index itself. No one can suppose that looking at the index can in any way interfere with the transfer of electricity or any other power from the hand to the board under it or to the table. If the board tends to move, it may do so, the index does not confine it; and if the table tends to move, there is no reason why it should not. If both were influenced by any power to move together they may do so,—as they did indeed when the apparatus was tied, and the mind and muscles left unwatched and unchecked.

I must bring this long description to a close. I am a little ashamed of it, for I think, in the present age, and in this part of the world, it ought not to have been required. Nevertheless, I hope it may be useful. There are many whom I do not expect to convince; but I may be allowed to say that I cannot undertake to answer such objections as may be made. I state my own convictions as an experimental philosopher, and find it no more necessary to enter into controversy on this point than on any other in science, as the nature of matter, or inertia, or the magnetization of light, on which I may differ from others. The world will decide sooner or later in all such cases, and I have no doubt very soon and correctly in the present instance. Those who may wish to see the particular construction of the test apparatus which I have employed, may have the opportunity at Mr. Newman's, 122, Regent Street. Further, I may say, I have sought earnestly for cases of lifting by attraction, and indications of attraction in any form, but have gained no traces of such effects. Finally, I beg to direct attention to the discourse delivered by Dr. Carpenter at the Royal Institution on the 12th of March 1852, entitled, 'On the influence of Suggestion in modifying and directing Muscular Movement, independently of Volition':—which, especially in the latter part, should be considered in reference to table-moving by all who are interested in the subject. M. FARADAY.

Royal Institution, June 27.

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[19]

A PK TEST WITH ELECTRONIC EQUIPMENT

By HELMUT SCHMIDT

ABSTRACT: The subjects in this research were tested for their psychokinetic ability by means of an electronic apparatus made up of a random number generator (RNG) connected with a display panel. The RNG produced random sequences of two numbers which were determined by a simple quantum process (the decay of radioactive strontium-90 nuclei). The essential aspect of the display panel was a circle of nine lamps which lighted one at a time in the clockwise (+1) direction or the counterclockwise (−1) direction depending on which of the two numbers the RNG produced. The subject's task was to choose either the clockwise or counterclockwise motion and try by PK to make the light proceed in that direction.

One run was made up of 128 "jumps" of the light, and there were four runs per session. In a preliminary series of 216 runs, the 18 subjects had a negative deviation of 129 hits. Accordingly, the main series was expected to give negative scores, and a negative attitude was encouraged among the subjects. Fifteen subjects carried out 256 runs, with a significant negative deviation of 302 hits ($P = .001$).

The RNG was checked for randomness throughout the experiment and was found to be adequate.—Ed.

In previous work (4, 5) the author was able to get significant evidence of precognition in which the testing apparatus was an electronic device based on a simple quantum process. The present experiment was an attempt to get significant evidence of psychokinesis by the use of a similar apparatus.

The basic part of the apparatus was a binary random number generator which produced the numbers "+1" and "−1" in random sequence, and the general objective was to have the subjects try to mentally influence the generator to produce one of the two numbers more frequently than the other.

The most easily available random generators, which have been used in many PK experiments, are a rolled die and a flipped coin. In comparison with these, an electronic random generator, the operation of which most of the subjects cannot understand, may at first thought seem psychologically unfavorable. Results of experiments with complex targets (3, p. 142), however, suggest that PK is goal oriented in the sense that results can be obtained by concentrating on the goal only, no matter how complicated the intermediate steps may seem to the rationalizing mind. A definite advantage of an electronic apparatus is that it permits a psychologically challenging formulation of the goal. In the present experiment the random number generator (RNG) was connected with a display panel showing a circle of nine lamps. One lamp was lit at a time, and each generated "+1" or "−1" caused the light to jump one step in the clockwise or counterclockwise direction, respectively. The subjects were not asked to try to force the generator to produce more +1's than −1's but, rather, to force the light on the panel to make more jumps in one direction or the other. Both tasks are certainly equivalent, but the latter seems psychologically much more appealing to most subjects.

A further obvious advantage of electronic test equipment is that the detailed results can be automatically recorded and evaluated and that one can work, if desired, at high speeds.

The particular type of random generator used here was chosen partly for practical and partly for theoretical reasons. The sequence in which the random numbers are produced is determined by simple quantum processes, the decays of radioactive strontium-90 nuclei. The electrons emitted in this decay trigger a Geiger counter, and the random times at which electrons are registered at the Geiger counter decide the generated numbers. Practically, the generator is easy to build, and the randomness of the generated numbers has been found to be very good. Furthermore, the simplicity of the generator allows a complete theoretical discussion (6) of its randomness properties; and in addition, one can say fairly well at which point the random element in the number generation comes in. The generator is essentially deterministic except for the random decay times of the nuclei.

The use of simple quantum jumps to provide randomness is, for

the theorist, a rather natural choice, since these processes are assumed by physicists to be nature's most elementary source of randomness, and some psi tests utilizing quantum processes have already been reported (**1, 2**). Certainly, the outcome of a die throw is also largely determined by microscopic quantum processes. The thermal vibrations of the surface and the air fluctuations at an atomic level co-determine the generated die face. The process in this case is much more complicated, however, since many more factors contribute to the end result.

Apparatus

The test equipment consisted of a binary random number generator and a display panel.

Random Number Generator

The RNG, which was similar to the one described in connection with earlier precognition experiments (**4, 5, 6**), can produce sequences of binary random numbers of any specified length. Electrons emitted by the strontium-90 decay trigger a Geiger counter and the momentary position of a binary high frequency counter at the time of the electron registration determines whether a "+1" or a "−1" is generated.

The numbers of electrical pulses produced on the +1 output and the −1 output are recorded by two electromechanical reset counters, and the complete sequence of generated numbers is recorded on paper punch tape.

Randomness Tests

Because of the simplicity of the circuitry, the degree of randomness to be expected of the RNG can be discussed in detail (**6**) and it can be shown to be much greater than required by the experiment.

The electronic circuitry is designed so that variations in the characteristics of the components cannot impair the randomness. In order to guard against any gross malfunctions, the proper electronic operation was tested frequently. Furthermore, the randomness of the generated number sequence was tested experimentally. For this purpose, a sequence of four million numbers, generated on many

different days, was recorded on paper tape. Then for the whole sequence the numbers N^+, N^- of generated +1's and −1's were counted and were found to be consistent with randomness, as was the total number of flips (F); i.e., events where a +1 was followed by a −1 or vice versa. The same procedure was applied to the 400 number sequences obtained by cutting the whole sequence into blocks of 10,000 numbers each. A goodness-of-fit test verified that the 400 values for +1 and −1, and the 400 values for F were consistent with their expected normal distribution.

Display Panel

In testing with this apparatus, the two above-mentioned counters for the numbers of generated +1's and −1's could serve as the only display, i.e., the (visual) focusing point toward which the subject could have directed his PK efforts. In this case, the subject might try to enforce mentally on the +1 counter a higher number of counts than on the −1 counter. It seemed desirable, however, to use a psychologically more stimulating display in the form of a panel with nine lamps arranged in a circle and connected to the RNG by a 30-foot long cable. One of the nine lamps was lighted at a time; and each time the RNG produced a signal, the light advanced one step in the clockwise or the counterclockwise direction according to whether the signal came to the +1 or the −1 output. Thus the light performed a "random walk" among the nine lamps. Rather than direct his PK toward the counters, then, the subject generally tried to "will" the light on the display panel to advance in an overall clockwise motion.

Some of the subjects, however, preferred to force the light in the counterclockwise direction. For them the two signal wires from the RNG to the display panel were interchanged by flipping a switch on the display panel so that a count on the +1 counter was displayed as a jump of the light in the counterclockwise direction. Thus, for all subjects, a jump of the lamp in the preferred direction, whether clockwise or counterclockwise, was registered on the +1 counter.

Test Procedure

The subjects in this experiment were members of the Institute for Parapsychology plus a few visitors. During a test session, the

subject sat in a dark closet with the display panel in front of him. The RNG and the experimenter were stationed in the room outside the closet.

Each testing session comprised 4 runs of 128 counts (steps in the random walk). A run took approximately two minutes. The machine stopped automatically after the one hundred and twenty-eighth count. There were short breaks, mostly between one-half and two minutes, between the runs.

At the beginning of each run, the subject, having decided in which direction (clockwise or counterclockwise) he wanted to influence the light to go, set a switch on the panel accordingly. Then the experimenter turned on the start switch, causing the RNG to generate 128 random numbers. At the end of each run, the experimenter recorded the readings of the +1 counter and −1 counter. The correctness of the counter readings was later checked with the sequence of generated numbers recorded on the paper punch tape.

From the experimenter's point of view, the subject's goal was always to produce a high number of +1 counts. From the subject's viewpoint the equivalent goal was to influence the light in the direction desired and indicated by the position of the switch on the display panel.

The subject was permitted to flip the switch during the course of a run so as to change the direction in which he wanted the hits displayed, but only a few subjects actually took advantage of the opportunity. With this arrangement, the subject could have had the impression that he was doing a test in precognition (by setting the switch in the direction in which he thought the light would move on the next jump) while he was actually doing a PK test.

Preliminary Series

There were 18 subjects in the preliminary series and they carried out a total of 54 sessions, each subject contributing from one to seven sessions.

The total score was a negative deviation of 129 hits below chance expectation out of 216 runs; $CR = 1.55$. These results include one subject who obtained a high positive score of 52 hits above chance in 16 runs ($CR = 2.3$).

Main Series

It was expected on the basis of the preliminary results that by leaving out the one high-scoring subject, an overall significantly negative score would be obtained in the main series. In order to emphasize the negative scoring, some subjects were asked to associate feelings of pessimism and failure with the experiment. The more negative-scoring subjects were used more frequently, and a few new subjects were allowed to contribute only after preliminary tests had suggested a negative scoring tendency.

The total length of the experiment was set in advance at 64 sessions of four runs each. It was not determined in advance, however, how many sessions each individual subject should contribute. Altogether there were 15 subjects and they contributed between one and 10 sessions each.

Although the proper randomness of the generator had been tested extensively, as mentioned before, a further safeguard against a possible bias of the generator was introduced. After the first half of the confirmatory test was completed the two outputs of the generator were internally interchanged. Thus, even a constant bias in the generator could not have caused the total significant score to be reported.

A total of 256 runs in this part of the experiment yielded a negative deviation of 302 hits ($CR = 3.33$; $P < .001$, two-tailed). Of the 64 sessions, 46 gave below-chance scores, 15 above-chance scores, and three were just at chance level ($CR = 4.0$). Of the 256 runs, 147 were below chance, 92 above chance, and 17 at chance level ($CR = 3.55$). These three CR values are certainly not independent, but they do emphasize the consistency of the results.

A post hoc analysis of the data showed two types of decline effect: more negative scoring in the second half of each run than in the first; and more negative scoring in the second half (the third and fourth runs) of each session than in the first half (first two runs). The decline results, however, are suggestive rather than statistically significant:

Deviation for pooled first half of the run: −91
Deviation for pooled second half of the run: −211
Deviation for the pooled first half of the session: −83
Deviation for the pooled second half of the session: −219

Discussion

The result of the experiment shows that the binary random number generator had no bias for generation of +1's or −1's as long as it was left unattended (in the randomness tests) but that it displayed a significant bias when the test subjects concentrated on the display panel, wishing for an increased generation rate of one number.

The experiment has been discussed in terms of PK, but in principle the result could certainly also be ascribed to precognition on the part of the experimenter or the subject. Since the sequence of generated numbers depended critically on the time when the test run began, and since the experimenter, in consensus with the subject, decided when to flip the start switch, precognition might have prompted experimenter and subject to start the run at a time which favored scoring in a certain direction.

If the PK interpretation is appropriate, the results imply the action of PK at some distance, since the generator was separated from the subject by a wall and only the display panel was close to the subject.

References

1. Beloff, J., and Evans, L. A radioactivity test of psychokinesis. *J. Soc. psych. Res.*, 1961, **41**, 41-46.
2. Chauvin, R., and Genthon, J. Eine Untersuchung über die Möglichkeit psychokinetischer Experimente mit Uranium und Geigerzähler [An investigation of the possibility of PK experiments with uranium and a Geiger counter]. *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie*, 1965, **8**, 140-47.
3. Rhine, L. E. *Mind Over Matter.* New York: Macmillan, 1970.
4. Schmidt, H. Anomalous prediction of quantum processes by some human subjects. Document D1-82-0821 (1969). Boeing Scientific Research Laboratories, Seattle, Washington. Available from Clearing House, U.S. Dept. Commerce, Springfield, Va.
5. Schmidt, H. Precognition of a quantum process. *J. Parapsychol.*, 1969, **33**, 99-108.
6. Schmidt, H. Quantum-mechanical random-number generator. *J. app. Physics*, 1970, **41** (No. 2), 462-68.

Institute for Parapsychology
College Station
Durham, N. C. 27708

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[20]

Evidence for Consciousness-Related Anomalies in Random Physical Systems

Dean I. Radin[1] **and Roger D. Nelson**[2]

Received May 6, 1988; revised June 12, 1989

Speculations about the role of consciousness in physical systems are frequently observed in the literature concerned with the interpretation of quantum mechanics. While only three experimental investigations can be found on this topic in physics journals, more than 800 relevant experiments have been reported in the literature of parapsychology. A well-defined body of empirical evidence from this domain was reviewed using meta-analytic techniques to assess methodological quality and overall effect size. Results showed effects conforming to chance expectation in control conditions and unequivocal non-chance effects in experimental conditions. This quantitative literature review agrees with the findings of two earlier reviews, suggesting the existence of some form of consciousness-related anomaly in random physical systems.

1. INTRODUCTION

The nature of the relationship between human consciousness and the physical world has intrigued philosophers for millenia. In this century, speculations about mind–body interactions persist, often contributed by physicists in discussions of the measurement problem in quantum mechanics. Virtually all of the founders of quantum theory—Planck, de Broglie, Heisenberg, Schrödinger, Einstein—considered this subject in depth,[1] and contemporary physicists continue this tradition.[2-7]

[1] Department of Psychology, Princeton University, Princeton, New Jersey 08544. Present address: Contel Technology Center, 15000 Conference Center Drive, P.O. Box 10814, Chantilly, Virginia 22021-3808.

[2] Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, New Jersey 08544.

The following expression of the problem can be found in a recent interpretation of quantum theory:

> If conscious choice can decide what particular observation I measure, and therefore into what states my consciousness splits, might not conscious choice also be able to influence the outcome of the measurement? One possible place where mind may influence matter is in quantum effects. Experiments on whether it is possible to affect the decay rates of nuclei by thinking suitable thoughts would presumably be easy to perform, and might be worth doing.[8]

Given the distinguished history of speculations about the role of consciousness in quantum mechanics, one might expect that the physics literature would contain a sizable body of empirical data on this topic. A search, however, reveals only three studies.

The first is in an article by Hall, Kim, McElroy, and Shimony, who reported an experiment "based upon taking seriously the proposal that the reduction of the wave packet is due to a mind–body interaction, in which both of the interacting systems are changed."[9] This experiment examined whether one person could detect if another person had previously observed a quantum mechanical event (gamma emission from sodium-22 atoms). The idea was based on the supposition that if person A's observation actually changes the physical state of a system, then when person B observes the same system later, B's experience may be different according to whether A has or has not looked at the system. Hall *et al.*'s results, based on a total of 554 trials, did not support the hypothesis; the observed number of "hits" obtained in their experiment was precisely the number expected by chance (277), while the variance of their measurements was significantly smaller than expected ($p < 0.05$).[9]

The second study is referred to by Hall *et al.*, who end their article by pointing out that a similar, unpublished experiment using cobalt-57 as the source was successful (40 hits out of 67 trials).[10]

The third study is a more systematic investigation reported by Jahn and Dunne,[11] who summarize results of over 25 million binary trials collected during seven years of experimentation with random-event generators. These experiments, involving long-term data collection with 33 unselected individuals, provide persuasive, replicable evidence of an anomalous correlation between conscious intention and the output of random number generators.

Thus, of three pertinent experiments referenced in mainstream physics journals, one describes results statistically too close to chance expectation and two describe positive effects.[9 11] Given the theoretical implications of such an effect, it is remarkable that no further experiments of this type can be found in the physics literature; but this is not to say that no such experiments have been performed. In fact, dozens of researchers have

reported conceptually identical experiments in the puzzling and uncertain domain of parapsychology. Perhaps because of the insular nature of scientific disciplines, the vast majority of these experiments are unknown to most scientists. A few critics who have considered this literature have dismissed the experiments as being flawed, nonreplicable, or open to fraud,[12 16] but their assertions are countered by at least two detailed reviews which provide strong statistical support for the existence of anomalous consciousness-related effects with random number generators.[17,18] In this paper, we describe the results of a comprehensive, quantitative meta-analysis which focused on the questions of methodological quality and replicability in these experiments.

2. THE EXPERIMENTS

The experiments involved some form of microelectronic random number generator (RNG), a human observer, and a set of instructions for the observer to attempt to "influence" the RNG to generate particular numbers, or changes in a distribution, solely by intention. RNGs are usually based upon a source of truly random events such as electronic noise, radioactive decay, or randomly seeded pseudorandom sequences.[19] Feedback about the distribution of random events is often provided in the form of a digital display, but audio feedback, computer graphics, and a variety of other mechanisms have also been used. Some of the RNGs described in the literature are technically sophisticated, the best devices employing electromagnetic shielding, environmental failsafe mechanisms triggered by deviant voltages, currents, or temperature, automatic computer-based data recording on magnetic media, redundant hard copy output, periodic randomness calibrations, and so on.[18,20]

RNGs are typically designed to produce a sequence of random bits at the press of a button. After generating a sequence of say, 100 random bits (0's or 1's), the number of 1's in the sequence may be provided as feedback. In an experimental protocol using a binary RNG, a run might consist of an observer being asked to cause the RNG to produce, in three successive button presses, a high number (sum of 1's greater than chance expectation of 50), a low number (less than 50), and a control condition with no directional intention. An experiment might consist of a group of individuals each contributing a hundred such runs, or one individual contributing several thousand runs. Results are usually analyzed by comparing high aim and low aim means against a control mean or theoretical chance expectation.

3. META-ANALYTIC PROCEDURES

The quantitative literature review, also called meta-analysis, has become a valuable tool in the behavioral and social sciences.[21] Meta-analysis is analogous to well-established procedures used in the physical sciences to determine parameters and constants. The technique assesses replication of an effect within a body of studies by examining the distribution of effect sizes.[22-24] In the present context, the null hypothesis (no mental influence on the RNG output) specifies an expected mean effect size of zero. A homogeneous distribution of effect sizes with nonzero mean indicates replication of an effect, and the size of the deviation of the mean from its expected value estimates the magnitude of the effect.

Meta-analyses assume that effects being compared are similar across different experiments, that is, that all studies seek to estimate the same population parameters. Thus the scope of a quantitative review must be strictly delimited to ensure appropriate commonality across the different studies that are combined.[21,25] This can present a nontrivial problem in meta-analytic reviews because replication studies typically investigate a number of variables in addition to those studied in the original experiments. In the present case, because different subjects, experimental protocols, and RNGs were employed within the reviewed literature, some heterogeneity attributable to these factors was expected in the obtained distribution of effect sizes. However, the circumscription for the review required that every study in the database have the same primary goal or hypothesis, and hence estimate the same underlying effect.

Experiments selected for review examined the following hypothesis: The statistical output of an electronic RNG is correlated with observer intention in accordance with prespecified instructions, as indicated by the directional shift of distribution parameters (usually the mean) from expected values.

Because this "directional shift" is most often reported as a standard normal deviate (i.e., Z score) in the reviewed experiments, we determined effect size as a Z score normalized by the square root of the sample size (N), $e = Z/\sqrt{N}$, where N was the total number of individual random events (with probability of a hit at $p = 0.5$, $p = 0.25$, etc.). This effect size measure is equivalent to a Pearson product moment correlation.[21]

3.1. Unit of Analysis

To avoid redundant inclusion of data in a meta-analysis, "units of analysis" are often specified. We employed the following method: If an author distinguished among several experiments reported in a single

article with titles such as "pilot test" or "confirmatory test," or provided independent statistical summaries, each of these studies was coded and quality-assessed separately. If an experiment consisted of two or more conditions comparing different intentions or types of RNG devices, the data were split into separate units of analysis to allow the results to be coded unambiguously. In general, within a given reviewed report, the largest possible aggregation of nonoverlapping data collected under a single intentional aim was defined as the unit of analysis (hereafter called an experiment or study).

For each experiment, a Z score was assigned corresponding to whether the observed result matched the direction of intention. Thus, a negative Z obtained under intention to "aim low" was recorded as a positive score. When sufficient data were provided in a report, Z was calculated from those data and compared with the reported results; the new calculation was used if there was a discrepancy. If only probability levels were reported, these were transformed into the corresponding Z score. For experiments reported only as "nonsignificant," a conservative value of $Z = 0$ was assigned; if the outcome was reported only as "statistically significant," $Z = 1.645$ was assigned; and if sample size was not reported or could not be calculated from the information provided, a special code of $N = 1$ was assigned.

3.2. Assessing Quality

Because the hypothesized anomalous effect is not easily accommodated within the prevailing scientific world-view, it is particularly important to assess the trustworthiness of each reviewed experiment. Unfortunately, estimating experimental quality tends to be a subjective task confounded by prior expectations and beliefs.[26,27] Estimates of inter-judge reliability in assessing the quality of research reports, for example, rarely exceed correlations of 0.5.[28] We addressed this problem by assigning to each experiment a single quality weight derived from a set of sixteen binary (present/absent) criteria. The first author coded and double-checked the coding for all studies; the second author independently coded the first 100 studies. Inter-judge reliability for quality criteria was $r = 0.802$ with 98 degrees of freedom.

These criteria were developed from published criticisms about random-number generator experiments[14,15,29–33] and from expert opinion on important methodological considerations when performing studies involving human behavior.[20,34,35] Collectively, these criteria form a measure of credibility by which to judge the reported data. The criteria assess the integrity of the experiment in four categories—procedures,

statistics, the data, and the RNG device—and they cover virtually all methodological criticisms raised to date. They are (1) control tests noted, (2) local controls conducted, (3) global controls conducted, (4) controls established through the experimental protocol, (5) randomness calibrations conducted, (6) failsafe equipment employed, (7) data automatically recorded, (8) redundant data recording employed, (9) data double checked, (10) data permanently archived, (11) targets alternated on successive trials, (12) data selection prevented by protocol or equipment, (13) fixed run lengths specified, (14) formal experiment declared, (15) tamper-resistant RNG employed, and (16) use of unselected subjects.

Each criterion was coded as being present or absent in the report of an experiment, specifically excluding consideration of previously published descriptions of RNG devices or control tests. This strategy was employed to reflect lower confidence in such experiments since, for example, randomness tests conducted once on an RNG do not guarantee acceptable performance in the same RNG in all future experiments. As a result, assessed quality was conservative, that is, lower than the "true" quality for some experiments, especially those reported only as abstracts or conference proceedings. Using unit weights (which have been shown to be robust in such applications[36]) on each of the sixteen descriptors, the quality rating for an individual experiment was simply the sum of the descriptors. Thus, while a quality score near zero indicated a low quality or poorly reported experiment, a score near sixteen reflected a highly credible experiment.

3.3. Assessing Effect Size

Assume that each of K experiments produces effect size estimates e of a parameter E, based on N samples, and that each e has a known standard error s. The weighted mean effect size is calculated as $e_\cdot = \sum \omega_i e_i / \sum \omega_i$, where $\omega_i = 1/s_i^2 = N_i$, and i ranges from 1 to K. The standard error of $e_\cdot$ is $s_e = (\sum \omega_i)^{-1/2}$. A test for homogeneity for the K estimates of e_i is given by $H_K = \sum \omega_i (e_i - e_\cdot)^2$, where H_K has a chi-square distribution with $K-1$ degrees of freedom.[37] The same procedure can be followed to test for homogeneity of effect size across M independent investigators. In this case, $e_{\cdot j}$ and s_{ej} are calculated per investigator, and the test for homogeneity is performed as $H_M = \sum \omega_j (e_{\cdot j} - e_{\cdot M})^2$, where $e_{\cdot j}$ and ω_j are mean weighted effect size and $1/s_e^2$ per investigator, respectively, $e_{\cdot M} = \sum \omega_j e_{\cdot j} / \sum \omega_j$, and j ranges from 1 to M. H_M has $M-1$ degrees of freedom.

For a quality-weighted analysis, we may determine $e_{\cdot Q} = \sum (Q_i \omega_i e_i) / \sum (Q_i \omega_i)$, where Q_i is the quality assessed for experiment i. The standard error associated with e_Q is $se_Q = (\sum (Q_i^2 \omega_i) / (\sum Q_i \omega_i)^2)^{-1/2}$; the test for homogeneity is similar to that described above. Finally, following

the practice of reviewers in the physical sciences,[23,24] we deleted potential "outlier" studies to obtain a homogeneous distribution of effect sizes and to reduce the possibility that the calculated mean effect size may have been spuriously enlarged by extreme values. The procedure used was as follows: If the homogeneity statistic for all studies was significant (at the $p < 0.05$ level), the study that would produce the largest reduction in this statistic was deleted; this was repeated until the homogeneity statistic had become nonsignificant.

4. RESULTS

On-line bibliographic databases for psychology and physics journals were searched, as was a specialized database covering parapsychological articles, technical reports, conference proceedings and manuscripts. Altogether 152 references were found from 1959 to 1987. These reports described 832 studies conducted by 68 different investigators (597 experimental studies and 235 control studies). Fifty-four experimental and 33 control studies reported only as nonsignificant were assigned $Z = 0$. Six experiments and two control studies coded as $(N = 1, Z > 0)$ were eliminated from further meta-analysis because effect size could not be accurately estimated (this required the elimination of one investigator who reported a single study). Figures 1 and 2 show the distributions of Z scores reported for control and experimental studies, respectively.

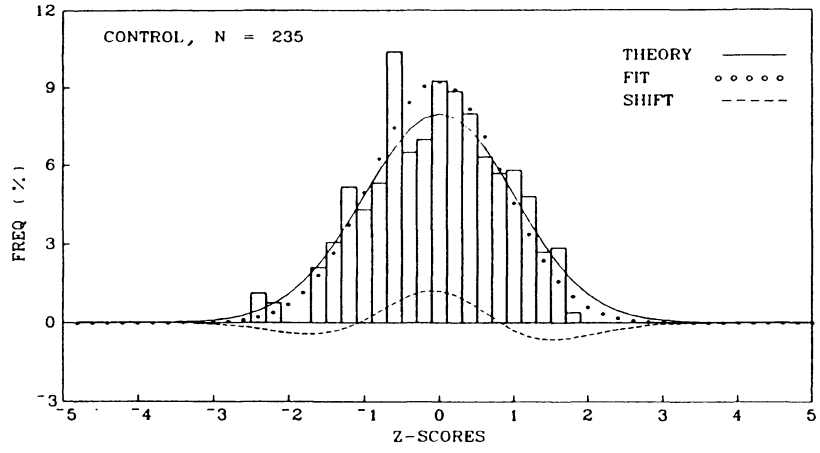

Fig. 1. Distribution of Z scores reported in 235 control studies. Thirty-three of these studies were reported only as "nonsignificant" and were assigned Z scores of zero. To replace the spurious spike at $Z = 0$, those 33 studies were recast as normally distributed Z scores, bounded by ± 1.64, averaging $Z = 0$.

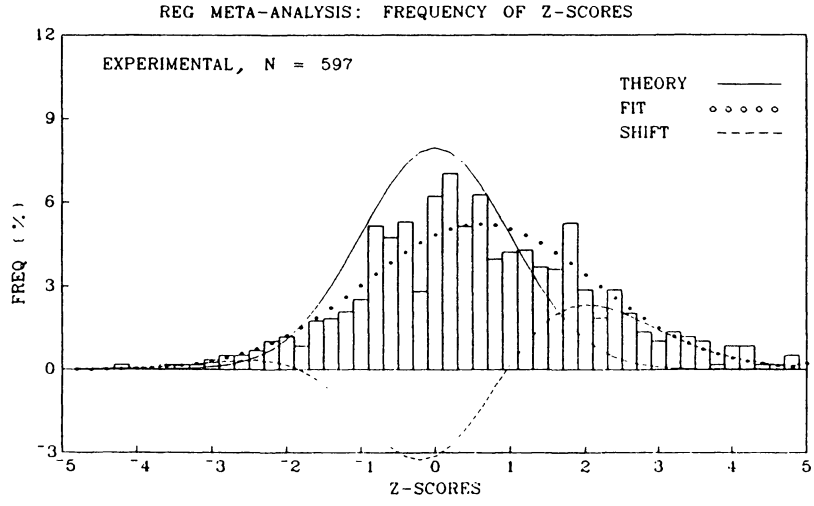

Fig. 2. Distribution of Z scores reported in 597 experimental studies. Fifty-four of these studies were reported as "nonsignificant" and were assigned Z scores of zero. As in Fig. 1, those 54 studies were recast as normally distributed Z scores, bounded by ± 1.64, averaging $Z = 0$.

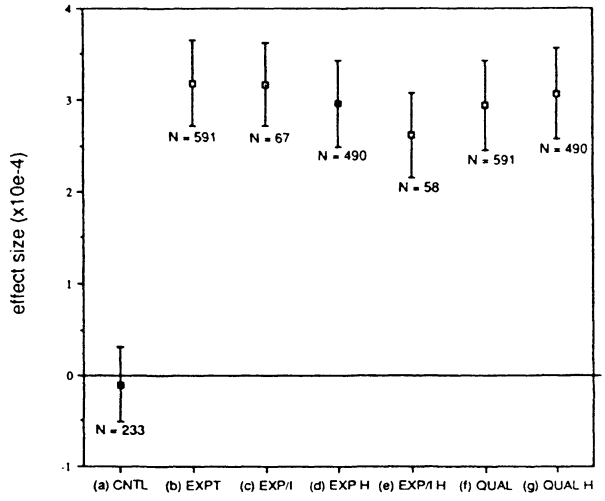

Fig. 3. Mean effect size point estimates ± 1 standard error for (a) control studies and (b) individual experiments; (c) mean effect size per investigator, (d) homogeneous mean effect size for experiments, (e) homogeneous mean effect size per investigator, (f) mean effect size for quality-weighted experiments, and (g) mean effect size for homogeneous quality-weighted experiments.

These results, expressed as overall mean effect sizes, show that control studies conform well to chance expectation (Fig. 3a), and that experimental effects, whether calculated for studies or investigators, deviate significantly from chance expectation (Fig. 3b, 3c). To obtain a homogeneous distribution of effect sizes, it was necessary to delete 17% of individual outlier studies (Fig. 3d) and 13% of mean effect sizes across investigators (Fig. 3e). This may be compared with exemplary physical and social science reviews, where it is sometimes necessary to discard as many as 45% of the studies to achieve a homogeneous effect size distribution.[19] Of individual studies deleted, 77% deviated from the overall mean in the positive direction, and of investigator means deleted, all were positive (i.e., supportive of the experimental hypothesis).

4.1. Effect of Quality

Some critics have postulated that as experimental quality increases in these studies, effect size would decrease, ultimately regressing to the "true" value of zero, i.e., chance results.[12,13,15,32,33,38] We tested this conjecture with two linear regressions of mean effect size vs. mean quality assessed per investigator, one weighted with ω_j as defined above and the other weighted with the number of studies per investigator. The calculated slope for the former is $-2.5 \times 10^{-5} \pm 3.2 \times 10^{-5}$, and for the latter, $-7.6 \times 10^{-4} \pm 3.9 \times 10^{-4}$. These nonsignificant relationships between quality and effect size is typical of meta-analytic findings in other fields,[39,40] suggesting that the present database is not compromised by poor experimental methodology. Another assessment of the effect of quality was obtained by comparing unweighted and quality-weighted effect sizes per experiment (Fig. 3b vs. 3f). These are nearly identical, and the same is true after deleting outliers to obtain a homogeneous quality-weighted distribution (Fig. 3d vs. 3g), confirming that differences in methodological quality are not significant predictors of effect size.

It might be argued that the quality assessment procedure employed here was nonoptimal because some quality criteria are more important than others, so that if appropriate weights were assigned, the quality-weighted effect size might turn out to be quite different. This was tested by Monte Carlo simulation, using sets of 16 weights, one per criterion, randomly selected over the range 0 to 6. A quality-weighted effect size was calculated for the 597 experiments as before, now using the random weights instead of unit weights, and this process was repeated one thousand times, yielding a distribution of possible quality ratings. The average effect size from the simulation was $3.18 \times 10^{-4} \pm 0.15 \times 10^{-4}$, indicating that in this particular database coded by these sixteen criteria,

the probable range of the quality-weighted mean effect size clearly excludes chance expectation of zero.

4.2. The "Filedrawer" Problem

Although accounting for differences in assessed quality does not nullify the effect, it is well known in the behavioral and social sciences that nonsignificant studies are published less often than significant studies (this is called the "filedrawer" problem[21,41-43]). If the number of nonsignificant studies in the filedrawer is large, this reporting bias may seriously inflate the effect size estimated in a meta-analysis. We explored several procedures for estimating the magnitude of this problem and to assess the possibility that the filedrawer problem can sufficiently explain the observed results.

The filedrawer hypothesis implicitly maintains that all or nearly all significant positive results are reported. If positive studies are not balanced by reports of studies having chance and negative outcomes, the empirical Z score distribution should show more than the expected proportion of scores in the positive tail beyond $Z = 1.645$. While no argument can be made that all negative effects are reported, it is interesting to note that the database contains 37 Z scores in the negative tail, where only 30 would be expected by chance. On the other hand, there are 152 scores in the positive tail, about five times as many as expected. The question is whether this excess represents a genuine deviation from the null hypothesis or a defect in reporting or editorial practices.

This question may be addressed by modeling based on the assumption that all significant positive results are reported. A four-parameter fit minimizing the chi-square goodness-of-fit statistic was applied to all observed data with $Z \geqslant 1.645$, using the exponential

$$Y = \frac{1}{\sqrt{2}\,\sigma} e^{-\sqrt{2}\,(|x-\mu|/\sigma)} \tag{1}$$

to simulate the effect of skew or kurtosis in producing the disproportionately long positive tail. This exponential is a probability distribution with the same mean and variance as the normal distribution, but with kurtosis = 3.0.

To begin, the null hypothesis of a (0, 1) normal distribution with no kurtosis was considered. To account for the excess in the positive tail, $N = 585{,}000$ filedrawer studies were required, and the chi-squared statistic remained far too large to indicate a reasonable fit (see Table I). This large N, in comparison with the 597 studies actually reported together with the poor goodness-of-fit statistic, suggests that the assumption of a (0, 1) normal distribution is inappropriate.

Table I. Four-Parameter Fit (E:N, N, Mean, sd) Minimizing Chi-Square (10 df) Goodness-of-Fit Statistic to the Positive Tail of the Observed Z Score Distribution, for Several Exponential:Normal Ratios[a]

Assumption	E:N ratio	N	Mean	sd	Chi-square	p
Normal distribution	0	585,000	0	1	57,867.84	0
(null hypothesis)	1	5,300	0	1	220.97	0
	2	4,800	0	1	167.84	0
	3	4,600	0	1	148.45	0
	10	4,400	0	1	119.69	0
Empirical distribution	0	700	0.145	2.10	23.94	0.008
	1	747	0.345	1.90	16.32	0.091
	2	757	0.445	1.80	14.21	0.164
	3	777	0.445	1.80	11.08	0.226
	10	807	0.445	1.80	11.08	0.351

[a] The null hypothesis is tested by clamping the mean at 0 and the standard deviation at 1, allowing N and E:N to vary. The empirical database is addressed by allowing all four parameters to vary.

Adding simulated kurtosis to a (0, 1) normal distribution by mixing exponential [Eq. (1)] and normal distributions in a 1:1 ratio reduced N by two orders of magnitude, and ratios of 2:1, 3:1, and 10:1 exponential to normal (E:N) yielded further small improvements. However, the chi-squared statistic still indicated a poor fit to the empirical data. Applying the same mixture of exponential and normal distributions, but starting from the observed values of $N = 597$, mean Z score $= 0.645$, and standard deviation $= 1.601$, with the constraint that the mean could only *decrease* from 0.645, resulted in much better fits to the data. Table I shows the results.

This procedure shows that the null hypothesis is unviable, even after allowing a huge filedrawer. The chi-square fit vastly improves with the addition of kurtosis, but only becomes a reasonably good fit when mean and standard deviation are allowed to approximate the empirical values. The filedrawer estimate from this model depends on a number of assumptions (e.g., the true distribution is generally normal, but has a disproportionately large positive tail). It suggests a total number of experimental studies on the order of 800, of which three-fourths have been formally reported.

A somewhat simpler modeling procedure was applied to the data assuming that all studies with significant Z scores in either the positive or negative tail are reported. The model is based on the normal distribution with a standard deviation $= 1$, and estimates the mean and N required to

account for the 152 Z scores in the positive tail and 37 Z scores in the negative tail. This mean-shift model, which ignores the shape of the observed distribution, results in an $N = 1{,}580$ and a mean Z score $= 0.34$.

These modeling efforts suggest that the number of unreported or unretrieved RNG studies falls in the range of 200 to 1,000. A remaining question is, how many filedrawer studies with an average *null* result would be required to reduce the effect to nonsignificance (i.e., $p < 0.05$)? This "failsafe" quantity is 54,000—approximately 90 times the number of studies actually reported. Rosenthal suggests that an effect can be considered robust if the failsafe number is more than five times the observed number of studies.[21]

5. DISCUSSION

Repeatable experiments are the keystone of experimental science. In practice, repeatability depends upon a host of controllable and uncontrollable ingredients, including factors such as stochastic variation, changes in environmental conditions, difficulties in communicating tacit knowledge employed by successful experimenters,[44] and so on. Difficulties in achieving systematic replication are therefore ubiquitous, from experimental psychology[21,45] to particle physics.[23,24] Of course, this is not to say that systematic replication is impossible in these or other fields, but it may appear to be extraordinarily difficult when experiments are considered individually rather than cumulatively. In the case of the present database, the authors of a recent report issued by the US National Research Council stated that the overall results of the RNG experiments could not be explained by chance,[46] but they questioned the quality and replicability of the research. This meta-analysis shows that effects are not a function of experimental quality, and that the replication rate is as good as that found in exemplary experiments in psychology and physics.

Besides the issue of replicability, five other objections are often raised about the present experiments. These are (a) the effect is inconsistent with prevailing scientific models, (b) the experimental methodology is technically naïve, thus the results are not trustworthy, (c) the experiments are vulnerable to fraud by subjects or by experimenters, (d) skeptics cannot obtain positive results, and (e) there are no adequate theoretical explanations or predictions for the anomalous effect.

These criticisms may be addressed as follows: (a) "Inconsistency with the scientific world-view" is essentially a philosophical argument that carries little weight in the face of repeatable experimental evidence, as suggested by the present and two corroborating meta-analyses.[17,18]

Indeed, if the "inconsistency" argument were sufficient to discount anomalous findings, we would have ignored much of the motivation leading to the development of quantum mechanics. (b) The "naïve methodology" argument was empirically addressed by the assessment of methodological quality in the present analysis. No significant relationship between quality and effect size was found. (c) Fraud postulated as the explanation of the results is untenable as it would have required widespread collusion among 68 independent investigators. In any case, even severe critics of parapsychological experiments have discounted fraud as a viable explanation.[32] (d) Skeptics often assert that only "believers" obtain positive results in such experiments. However, a thorough literature search finds not a single attempted replication of the RNG experiment by a publicly proclaimed skeptic; thus the assertion is not based on verifiable evidence. Furthermore, skeptics who claim to have attempted replications insist (without providing details or references) that they have never achieved positive results in any of their RNG experiments.[15,47] Such a claim is itself quite remarkable, as the likelihood of never obtaining a statistically significant result by chance in series of experiments can be extremely low, depending on the number of experiments conducted. Unfortunately, because we cannot determine how many experiments skeptics have actually conducted, it is impossible to judge the validity of this criticism.

Finally, (e) the "no theoretical basis" argument is correct, but it does not support a negative conclusion about experimental observation. There are at present no adequate theories, with the possible exception of some interpretations of quantum mechanics,[2,3,8,11] that convincingly explain or predict consciousness-related anomalies in random physical systems. We note, however, that the anomalous effects reviewed in this paper apparently can be operationally predicted under well-specified conditions. For example, when individuals are instructed to "aim" for high (or low) numbers in RNG experiments, it is possible to predict with some small degree of confidence that anomalous positive (or negative) shifts of distribution means will be observed.

6. CONCLUSION

In this paper, we have summarized results of all known experiments testing possible interactions between consciousness and the statistical behavior of random-number generators. The overall effect size obtained in experimental conditions cannot be adequately explained by methodological flaws or selective reporting practices. Therefore, after considering all of the

retrievable evidence, published and unpublished, tempered by all legitimate criticisms raised to date, it is difficult to avoid the conclusion that under certain circumstances, consciousness interacts with random physical systems. Whether this effect will ultimately be established as an overlooked methodological artifact, as a novel bioelectrical perturbation of sensitive electronic devices, or as an empirical contribution to the philosophy of mind, remains to be seen.

ACKNOWLEDGMENTS

This study was supported by major grants from the James S. McDonnell Foundation, Inc. and the John E. Fetzer Foundation, Inc. The authors express their gratitude to Dr. York Dobyns of the Princeton University Engineering Anomalies Laboratory for his assistance with the filedrawer models.

REFERENCES

1. R. G. Jahn and B. J. Dunne, *Margins of Reality* (Harcourt Brace Jovanovich, Orlando, Florida, 1987).
2. B. d'Espagnat, "The quantum theory and reality," *Sci. Am.*, pp. 158–181 (November, 1979).
3. O. Costa de Beauregard, "S-matrix, Feynman zigzag and Einstein correlation," *Phys. Lett.* **67A**, 171–173 (1978).
4. N. D. Mermin, "Is the moon there when nobody looks? Reality and the quantum theory," *Phys. Today*, pp. 38–47 (April, 1985).
5. A. Shimony, "Role of the observer in quantum theory," *Am. J. Phys.* **31**, 755 (1963).
6. E. P. Wigner, "The problem of measurement," *Am. J. Phys.* **31**, 6 (1963).
7. U. Ziemelis, "Quantum-mechanical reality, consciousness and creativity," *Can. Res.* **19**, 62–68 (September, 1986).
8. E. J. Squires, "Many views of one world—an interpretation of quantum theory," *Eur. J. Phys.* **8**, 173 (1987).
9. J. Hall, C. Kim, B. McElroy, and A. Shimony, "Wave-packet reduction as a medium of communication," *Found. Phys.* **7**, 759–767 (1977); p. 761.
10. R. Smith, unpublished manuscript, MIT, 1968. (Cited in Ref. 9, p. 767.)
11. R. G. Jahn and B. J. Dunne, "On the quantum mechanics of consciousness, with application to anomalous phenomena," *Found. Phys.* **16**, 721–772 (1986).
12. J. E. Alcock, *Parapsychology: Science or Magic?* (Pergamon Press, Elmsford, New York, 1981), pp. 124–125.
13. M. Gardner, *Science: Good, Bad, and Bogus* (Prometheus Books, Buffalo, New York, 1981).
14. R. Hyman, "Parapsychological research: A tutorial review and critical appraisal," *Proc. IEEE* **74**, 823–849 (1986).
15. P. Kurtz, "Is parapsychology a science?" In *Paranormal Borderlands of Science*, K. Frazier, ed. (Prometheus Books, Buffalo, New York, 1981).

16. D. F. Marks, "Investigating the paranormal," *Nature (London)* **320**, 119–124 (1986).
17. C. Honorton, "Replicability, experimenter influence, and parapsychology: An empirical context for the study of mind," paper presented at the annual meeting of the AAAS, Washington, D.C., 1978.
18. E. C. May, B. S. Humphrey, and G. S. Hubbard, "Electronic system perturbation techniques." SRI International Final Report, September 30, 1980.
19. H. Schmidt, "Precognition of a quantum process," *J. Parapsychol.* **33**, 99–108 (1969); "A PK test with electronic equipment," *J. Parapsychol.* **34**, 175–181 (1970); "Mental influence on random events," *New Sci. Sci. J.* **50**, 757–758 (1971); "PK tests with pre-recorded and pre-inspected seed numbers," *J. Parapsychol.* **45**, 87–98 (1981).
20. R. G. Jahn, "The persistent paradox of psychic phenomena: An engineering perspective," *Proc. IEEE* **70**, 136–170 (1982); R. D. Nelson, B. J. Dunne, and R. G. Jahn, "An REG experiment with large data-base capability, III: Operator-related anomalies," Technical Note PEAR 84003, Princeton Engineering Anomalies Research Laboratory, Princeton University, School of Engineering/Applied Science, September 1984; H. Schmidt, R. Morris, and L. Rudolph, "Channeling evidence for a PK effect to independent observers," *J. Parapsychol.* **50**, 1–16 (1986).
21. R. Rosenthal, *Meta-Analytic Procedures for Social Research* (Sage Publications, Beverly Hills, California, 1984); K. Wachter, "Disturbed by meta-analysis?" *Science* **241**, 1407–1408 (1988). We may note that Cohen's *h*, the difference between control and experimental proportions, is a common effect size measure that might have been used in the present study. This was rejected in favor of *e*, as defined, because some of the reviewed studies reported only final *p* values or only overall *Z* scores; *e* was thus deemed more useful in the present meta-analysis.
22. R. L. Bangert-Drowns, "Review of developments in meta-analytic method," *Psychol. Bull.* **99**, 388–399 (1986).
23. A. H. Rosenfeld, "The particle data group: Growth and operations." *Annu. Rev. Nucl. Sci.* **25**, 555–599 (1975).
24. C. G. Wohl *et al.*, *Rev. Mod. Phys.* **56**, Part II, p. S5 (1984).
25. G. V. Glass, "In defense of generalization," *Behav. Brain Sci.* **3**, 394–395 (1978).
26. H. M. Cooper, "Scientific guidelines for conducting integrative reviews," *Rev. Educ. Res.* **52**, 291–302 (1982).
27. R. M. Dawes, "You can't systematize human judgment: Dyslexia." In *New Directions for Methodology of Social and Behavioral Science: Fallible Judgment in Behavorial Research*, R. A. Shweder, ed. (Jossey-Bass, San Francisco, 1980), pp. 67–78.
28. S. D. Gottfredson, "Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgments," *Am. Psychol.* **33**, 920–934 (1978).
29. C. Akers, "Methodological criticisms of parapsychology." In *Advances in Parapsychological Research*, Vol. 4, S. Krippner, ed. (McFarland, Jefferson, North Carolina, 1984); "Can meta-analysis resolve the ESP controversy?" In *A Skeptic's Handbook of Parapsychology*, P. Kurtz, ed. (Prometheus Books, Buffalo, New York, 1985).
30. J. E. Alcock, "Parapsychology: Science of the anomalous or search for the soul," *Behav. Brain Sci.* **10**, 553–565 (1987).
31. P. Diaconis, "Statistical problems in ESP research," *Science* **201**, 131–136 (1978).
32. C. E. M. Hansel, *ESP and Parapsychology: A Critical Reevaluation* (Prometheus Books, Buffalo, New York, 1980).
33. R. Hyman, "The ganzfeld psi experiment: A critical appraisal," *J. Parapsychol.* **49**, 3–50 (1985).
34. T. X. Barber, *Pitfalls in Human Research: Ten Pivotal Points* (Pergamon Press, Elmsford, New York, 1976).

35. J. B. Rhine, "Comments: 'A new case of experimenter unreliability.'" *J. Parapsychol.* **38**, 215–255 (1974).
36. R. M. Dawes, "The robust beauty of improper linear models in decision making," *Am. Psychol.* **34**, 571–582 (1979).
37. L. V. Hedges, "How hard is hard science, how soft is soft science?" *Am. Psychol.* **42**, 443–455 (1987).
38. C. E. M. Hansel, *ESP: A Scientific Evaluation* (Charles Scribner's Sons, New York, 1966), p. 234.
39. R. Rosenthal and D. B. Rubin, "Interpersonal expectancy effects: The first 345 studies," *Behav. Brain Sci.* **3**, 377–415 (1978).
40. G. V. Glass, B. McGaw, and M. L. Smith, *Meta-analysis in Social Research* (Sage Publications, Beverly Hills, California, 1981).
41. Q. McNemar, "At random: Sense and nonsense," *Am. Psychol.* **15**, 295–300 (1960).
42. S. Iyengar and J. B. Greenhouse, "Selection models and the file-drawer problem," Technical Report 394, Department of Statistics, Carnegie-Mellon University (July, 1987).
43. L. V. Hedges, "Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences," *J. Educ. Stat.* **9**, 61–86 (1984).
44. H. H. Collins, *Changing Order: Replication and Induction in Scientific Practice* (Sage Publications, Beverly Hills, California, 1985).
45. S. Epstein, "The stability of behavior, II: Implications for psychological research," *Am. Psychol.* **35**, 790–806 (1980).
46. D. Druckman and J. A. Swets, eds. *Enhancing Human Performance: Issues, Theories, and Techniques* (National Academy Press, Washington, D.C., 1988), p. 207.
47. A. Neher, *The Psychology of Transcendence* (Prentice-Hall, Englewood Cliffs, New Jersey, 1980), p. 147.

[21]

FURTHER STUDIES OF AUTONOMIC DETECTION OF REMOTE STARING: REPLICATION, NEW CONTROL PROCEDURES, AND PERSONALITY CORRELATES

BY WILLIAM BRAUD, DONNA SHAFER, AND SPERRY ANDREWS

ABSTRACT: In a previous paper, we reviewed early experimental attempts to assess subjects' accuracy in consciously detecting when they are being watched or stared at by someone situated beyond the range of their conventional senses. We also reported new results of our own experiments in which a more "unconscious" autonomic nervous system reaction (spontaneous electrodermal activity) was used to assess accuracy of detection of staring (remote attention). In our experiments, one subject (the starer) directed full attention to another distant subject's (staree's) image on the monitor of a closed-circuit television system used to eliminate the possibility of subtle sensory cues. The staree's spontaneous electrodermal activity, meanwhile, was monitored objectively by a computer system during randomly interspersed staring and nonstaring periods; the staree was blind regarding the number, timing, and sequencing of the two types of period. We found evidence for significant blind autonomic discrimination between the staring and nonstaring episodes. In the present paper, we report evidence for autonomic discrimination of staring versus nonstaring periods in two replications—one involving the same starer who had participated in the earlier studies (t [15] = 2.08; p = .05, two-tailed; effect size r = .47), and the second involving three new starers (t [29] = 1.92; p = .06, two-tailed; effect size r = .34). Chance results were found, as expected, in a new, improved control condition (a "sham control") in which the data were treated as they were in a true staring study, but staring did not, in fact, occur. We also found that the magnitude of the remote autonomic staring detection effect was significantly related to the starees' degree of introversion (Myers-Briggs Type Indicator) and to their degree of social avoidance and distress (social anxiety).

In a previous paper (Braud, Shafer, & Andrews, 1993), we reviewed the scientific literature dealing with the purported ability to detect when one is being watched or stared at by someone situated beyond the range of the conventional senses. Surveys indicated that between 68% and 94% of various samples reported having had staring detection experiences in their everyday lives. Previous investigations provided suggestive evidence that persons were indeed able to detect, consciously, when they were being stared at under conditions in which precautions were taken to

This article is based on a paper presented at the Thirty-Fifth Annual Convention of the Parapsychological Association in Las Vegas, NV, in August, 1992.

Please address all correspondence to Dr. William Braud, Institute of Transpersonal Psychology, 744 San Antonio Road, Palo Alto, CA 94303.

eliminate possible subtle sensory cues. In particular, positive conscious-guessing results were obtained in two studies in which sensory cueing was eliminated through use of one-way mirrors (Peterson, 1978) and use of a closed-circuit television system (Williams, 1983).

We hypothesized that stronger effects might be obtained if relatively "unconscious" autonomic nervous system activity were used as the indicator of staring detection, rather than conscious guessing. Our reasoning was that autonomic reactions might be less distorted by higher cognitive processes and therefore might provide a purer and more sensitive indicator. We presented the results of two original experiments in which sympathetic nervous system activation was assessed by means of electrodermal monitoring during randomly interspersed remote-staring and nonstaring (control) periods. The monitored participant was unaware of the number, timing, or scheduling pattern of these two types of periods. The possibility of sensory cueing was eliminated through the use of a closed-circuit television system for staring: the starer devoted full attention to the staree's image on the television monitor. In the first experiment, 16 untrained participants evidenced significant autonomic discrimination, becoming more activated during staring than during nonstaring periods. In the second experiment, 16 subjects who had been extensively trained to become more aware of their interconnections with other people and less defensive about their connectedness also evidenced significant autonomic discrimination, but became more calm during staring than during nonstaring periods; the starer had been similarly trained. As judged by effect sizes, unconscious autonomic detection did indeed appear to yield stronger effects than did previous conscious verbal or motoric detection assessments.

In the present paper, we present our attempts to replicate and extend our previous findings. Identical equipment, basic procedures, and analysis methods were used. The first replication involved 3 new starers and 30 new starees. The second replication involved the same starer who had participated in the earlier experiments reported in 1993, but employed 16 new starees. We made two additions in these studies. One of these was the introduction (into Replication 2) of an additional empirical control condition. This was a "sham control" in which we treated sessions and data as we did for real staring sessions, but staring did not, in fact, occur; this provided an empirical assessment of the likelihood of obtaining chance discriminations of otherwise equivalent session segments. The second improvement was the introduction of a new personality assessment for the starees in both replications. In addition to the Myers-Briggs Type Inventory (MBTI) we had been using in the original studies, we

included an assessment of social anxiety or discomfort in a social situation (a Social Avoidance and Distress scale), in order to explore the possible interrelationships of these personality characteristics with the autonomic staring detection effect.

Method

Subjects

Thirty volunteer participants (22 females and 8 males) served as "starees" for Replication 1, and 16 volunteers (5 males and 11 females) participated as starees for Replication 2. In Replication 1, half of the starees were persons already known by the starers (relatives, friends, or familiar undergraduate classmates), whereas half were unknown at the time of the laboratory session (i.e., they were unfamiliar undergraduates); only one of the starees had participated previously in laboratory psi experiments. (Later results did not differ for the known versus unknown starees.) It had been decided in advance that each starer was to work with 10 starees and that results for all 30 starees were to be pooled for purposes of analysis. In Replication 2, 13 of the starees were previously unknown undergraduate students from a local college, and 3 were friends or relatives of the starer; only 2 of the starees had participated previously in laboratory psi studies. Participants were selected on the basis of availability during planned laboratory session times and on the basis of interest in participating in a study exploring the "feeling of being stared at." Across both replications, staree age ranged from 17 years to 40 years.

The starers of Replication 1 were three undergraduate psychology students (two females and one male) from a local college who were participating in independent studies internships at the Mind Science Foundation. None of these starers had prior laboratory psi research experience. The starers were trained for the experiment by the second author (D.S.), who had served as starer in our original (1993) staring detection experiments. D.S. served as starer for Replication 2. She herself had participated previously in extensive "connectedness" training that had been provided by the third author, S.A. This training (which is described in Braud, Shafer, & Andrews, 1993) took the form of approximately 20 hours of intellectual and experiential exercises designed to help individuals become more adept at and comfortable with experiencing interconnections with others, and to become more aware of, and to

deal more effectively with, psychological resistances to such connectedness. It is important to note that D.S. was very comfortable with "connecting with" (i.e., having feelings of "merging with") others when these replication experiments began, and that it is likely that she communicated this ease and comfort to the three Replication 1 starers during the course of their training by her. (This training involved discussions of the rationale for the studies, previous results, and procedural information about the experiments; the three Replication 1 starers experienced no formal "connectedness" training, although that training was discussed in general terms by D.S.)

Apparatus

The experimental apparatus was identical to that described in Braud, Shafer, and Andrews (1993) and consisted of silver/silver chloride palmar electrodes (7.0 mm in diameter) attached with semi-conductive electrode gel, a skin-resistance amplifier (Lafayette Model 76405), and an analog-to-digital converter interfaced with a microcomputer. A color video camera (Hitachi Camcorder VM-2250) in the staree's room permitted the staree to be viewed by the starer in a distant room without the possibility of sensory cueing. The camera's radio frequency output was boosted by a 10-dB amplifier, then conveyed via heavy duty 300-ohm impedance twin-lead cable to a 19-in. color TV monitor (Sony Trinitron KV-1914) situated in the starer's room. Additional details concerning equipment, room layout, and physiological monitoring are given in Braud and Schlitz (1989).

Procedure

With the exceptions to be noted later, procedural details for the two replications were identical to those of the original experiments (see Braud, Shafer, & Andrews, 1993). For both replications, the starer (who was also the experimenter) greeted the staree in the starer's room, explained the experiment, and showed the staree the television monitor on which the latter's image would appear during the session.

Next, the starer led the staree to the staree's room, which was located in an entirely different suite area across an outside corridor. The two rooms were separated from each other by two inner hallways, an outer corridor, and four closed doors. Neither room contained any windows. Conventional sensorimotor communication between these two rooms, under the conditions of the experiment, was not possible. The staree's

room was brightly illuminated by means of overhead fluorescent lights. The camera, which was active continuously throughout the entire session, was mounted on a tripod 6 ft away from the staree's chair, at eye level, and at an angle of approximately 45 degrees left of center (from the staree's point of view). The camera's zoom lens was set so that the staree's shoulders, neck, and head would be visible on the monitor in the starer's room. The camera's autofocus function was disabled in order to eliminate distracting camera lens movement noises that otherwise might have resulted from automatic tracking of staree movements; this also eliminated possible distracting changes in the staree's image, from the starer's point of view.

The staree was seated in a comfortable recliner chair (which remained in an upright position throughout the experiment), and the experimenter attached two silver/silver chloride electrodes filled with partially conductive gel to the staree's left palm by means of adhesive electrode collars. The staree was asked to sit quietly for the next 20 min and to refrain from unnecessary movements (especially of the left hand and arm). In order to more closely simulate a naturalistic staring-detection situation, the staree occupied his or her mind during the session by studying, reading a magazine, or thinking about and planning the day's activities (for Replication 1), or by completing the personality assessments (in Replication 2). The staree was told that the camera would be on throughout the 20-min session, but that the experimenter would watch the monitor only at certain randomly determined times. At those times, the starer would stare intently at the staree's image on the monitor and would attempt to gain the staree's attention. The staree was asked not to try to guess consciously when those periods (of which the staree was, of course, kept blind) might be occurring and was told that we were exploring whether any unconscious physiological reactions might be associated with remote staring. The experimenter then left the subject alone in the staree room and went to the distant starer's room, closing all intervening doors.

In the starer's room, the experimenter/starer recorded the staree's basal skin resistance and then, prior to starting the microcomputer that controlled the session events, retrieved from a hidden location a sealed opaque envelope that contained the random sequence of staring and nonstaring periods that would be used for that session. Forty-six such envelopes had been prepared previously by W. B., who had used a computer's random algorithm to generate the random sequence of the 10 staring and 10 nonstaring periods for each session. In a hidden location known only to him, W. B. kept his own copies of the 46 random

sequences. The microcomputer program controlled the timing of the various events of the experiment and recorded the staree's electrodermal activity during each of the twenty 30-sec recording periods. Each of the 20 recording periods was signaled by a low-pitched tone (audible only to the experimenter, through headphones); a 30-sec rest period followed each recording period. The experimenter/starer consulted the contents of the session envelope to learn which of the 20 recording epochs were to be devoted to staring and which were to serve as the nonstaring control periods. If the random sequence indicated a staring period the experimenter/starer silently swiveled his or her chair around so that it faced the television monitor, and stared intently at the staree's monitor image throughout the 30-sec recording periods. During nonstaring periods, the experimenter/starer kept the chair turned away from the monitor, so that the monitor's screen could not be seen, and thought about matters unrelated to the experiment. All reflective surfaces had been carefully covered and inadvertent glimpses of the monitor screen were not possible. In consulting a session's random-sequence sheet, the experimenter/starer used a method of occluding all epoch instructions other than the present one, so that he/she could devote full attention to the assignment for that epoch without being distracted by instructions for subsequent or previous epochs.

Throughout the session, the experimenter/starer was provided with no information about the staree's ongoing electrodermal activity; the latter was continuously and automatically assessed by the computer system. The equipment sampled the staree's rectified (by means of a diode) spontaneous phasic skin resistance responses (SRR) 10 times each second for the 30 seconds of a recording epoch and averaged these measures, providing what is virtually a measure of the area under the curve described by the fluctuation of electrodermal activity over time (i.e., the mathematically integrated activity). Because of the slowly changing nature of these autonomic reactions, this relatively slow sampling rate is quite adequate. At the end of the experimental session, the computer printed the electrodermal results for each of the 20 recording periods. The experimenter filed away the printout, taking special precautions not to look at the electrodermal measures, then went to the staree's room and discussed the experiment in general terms with the staree. Neither experimenter/starer nor staree had any knowledge of the numerical results for the session. Only after all sessions had been completed did W.B. analyze the results and give the experimenter feedback. The experimenter later provided feedback to those starees who requested it.

The general procedure described above applied to both replications. The specifics, however, were changed slightly for Replication 2. For the latter, instead of 20 recording epochs (10 staring and 10 nonstaring, randomly interspersed), a session consisted of 32 recording epochs. One half of a session was an "experimental" half and included 8 staring and 8 nonstaring periods (randomly interspersed). The other half of a session served as a new empirical sham control and included 8 sham or pseudostaring periods and 8 nonstaring periods (also randomly interspersed). The experimental half provided a comparison of true staring versus the absence of staring. The sham control half provided data that were treated in the same manner as the experimental data, but actual staring did not take place during the sham/pseudostaring periods. This provided an empirical control that yielded information about the likelihood of artifactual evidence for autonomic discrimination in arbitrary subdivisions of a session. For half of the Replication 2 sessions, the experimental half preceded the sham half; for the remaining sessions, the sham half preceded the experimental half. The television monitor was turned off throughout the sham half of a session, and the experimenter/starer occupied her mind with matters unrelated to the experiment. The staree had no way of knowing, conventionally, when the real periods and when the sham periods were in effect; and he/she was monitored in an identical fashion throughout both types of period.

Personality Assessments

For Replication 1, one personality assessment was administered. This was the Social Avoidance and Distress (SAD) scale (Watson & Friend, 1969) which measures social-evaluative anxiety (the experience of distress, discomfort, fear, and anxiety in social situations) and deliberate avoidance of social situations. This self-report scale emphasizes subjective experience, and it excludes physiological signs as well as items related to impaired performance. The scale is constructed so that the opposite instance of a trait simply indicates the absence of that trait, not the presence of some other trait. For example, the opposite instance of social avoidance is simply lack of an avoidance motive, not desire to affiliate. Similarly, the opposite instance of distress is lack of unhappiness, not the presence of some positive emotion. Others have found that scoring patterns of the SAD scale were indeed predictive of behaviors in social situations. We sought to learn whether SAD scoring might also be predictive of reactions to the remote or "psi mediated" social conditions involved in remote attention (remote staring or watching).

For Replication 2, the SAD scale was used along with the Myers-Briggs Type Indicator (MBTI, Form F: see Briggs & Myers, 1957). For this study, we were especially interested in the MBTI extraversion-introversion scale because of its possible relationships with SAD scoring and with remote-staring detection effects in this psi-mediated social (staring) context.

For Replication 1, the psychological assessments were completed by the starees after their experimental sessions. For Replication 2, the psychological assessments were completed by starees during their experimental sessions.

Experimental Hypotheses

Our experimental hypotheses were that, in Replications 1 and 2, the starees would discriminate the true staring from the nonstaring periods autonomically (electrodermally)—that their levels of spontaneous electrodermal activity during the staring periods would differ from those during the nonstaring periods. Therefore, two-tailed tests were used in the analyses, with alpha set at $\leq .05$. We also predicted that, in Replication 2, no such discrimination would occur in the empirical (sham) control segments of the sessions.

Exploratory analyses examined the correlations among the magnitude of the autonomic remote staring detection effect, SAD scoring, and MBTI extraversion-introversion scoring. Since these analyses were exploratory, two-tailed tests were used in their evaluation, with alpha set at $\leq .05$.

Results

Primary Analyses

For each volunteer participant (staree), electrodermal activity was measured during 10 staring and 10 nonstaring periods (for Replication 1) or during 8 staring and 8 nonstaring periods (for Replication 2). Rather than compare these multiple scores within a given participant, we reduced the activities for an entire session to a single score for each participant and performed statistical tests using participants, instead of multiple period scores, as the units of analysis. We used the more conservative session score (a kind of single majority-vote score) in order to bypass criticisms based on possible nonindependence of multiple electrodermal measures taken within a given session. Although it would

be possible to analyze individual epoch scores using, for example, a repeated measures analysis of variance procedure, such an analysis assumes that the autocorrelations among the measures within each session (i.e., within each participant) are constant across epochs, and that the same autocorrelation applies to all sessions (participants) (J. Utts, personal communication, July 13, 1991). Because these assumptions may not be met in these experiments, we preferred to use the more conservative session-based (rather than epoch-based) analyses, even though the former are more wasteful of data and result in tests with reduced statistical power.

For each of the 30 sessions for Replication 1, a total score was calculated for all 20 recording periods (10 staring and 10 nonstaring). This total score was divided into the sum of the electrodermal activity scores for the 10 staring (*S*) periods; the process was repeated for the 10 nonstaring (*N*) periods. In the absence of a remote-staring effect, these two ratios [*S/(S+N), N/(S+N)*] should approximate 50%. A remote staring effect would be indicated by a significant departure of the scores from the 50% mean chance expectation (MCE). A single mean *t* test was used to assess the departure of the ratios from MCE (50%). This is approximately equivalent to calculating a dependent (matched) *t* test to compare the raw scores for each subject for staring versus nonstaring periods. We have consistently used such ratio scores in our various projects as a method of "standardizing" scoring so that scoring magnitude could be compared more meaningfully for the different dependent measures (response systems) with which we work. A similar analysis was performed for the 16 sessions of Replication 2, with 8 staring and 8 nonstaring periods contributing to each session score.

For Replication 1, mean electrodermal activity percentages were 45.15% (for the staring periods) and 54.85% (for the nonstaring periods), rather than the 50%/50% expected by chance. This scoring rate yielded a single mean $t = 1.85$ (29), $p = .06$, two-tailed, and an effect size (r) = .34. For Replication 2, mean electrodermal activity percentages were 45.66% (for the staring periods) and 54.34% (for the nonstaring periods), rather than the 50%/50% expected by chance. This scoring rate yielded a single mean t (15) = 2.08; $p = .05$, two-tailed, and an effect size (r) = .47. Thus, for both replications, the autonomic discrimination took the form of reduced spontaneous electrodermal activity during staring periods, compared with nonstaring periods. For the 16 sham control session scores of Replication 2 (each based on 8 pseudostaring and 8 nonstaring periods), the mean electrodermal activity percentages were, as expected, virtually identical to the 50%/50% values expected on

the basis of chance. Here the scoring rates were 49.16% (for the pseudostaring periods) and 50.84% (for the nonstaring periods). This scoring rate yielded a single mean t (15) = 0.30; p = .76, two-tailed; and an effect size (r) = .08. Expanded summary statistics for Replications 1 and 2 and for the Sham Control series are presented in Table 1. For comparative purposes, the results for our previous two series with untrained and trained starees (see Braud, Shafer & Andrews, 1993) are also included in this table. Electrodermal activity rates during the staring and nonstaring periods of all four experiments, as well as for the sham control sessions, are presented graphically in Figure 1.

TABLE 1
STATISTICAL SUMMARY OF AUTONOMIC STARING DETECTION RESULTS FOR FOUR EXPERIMENTS AND FOR THE SHAM CONTROL SERIES

Series	Scoring rate $\overline{X}$	Scoring rate *SD*	Single mean *t*	*df*	*p*[a]	*z*[b]	Effect size[c] *r*	95% Confidence interval
Untrained Ss	59.38%	14.11	−2.66	15	.02	−2.37	−.57	51.86–66.90
Trained Ss	45.45%	8.46	2.15	15	.05	1.98	.48	40.94–49.95
Replication 1	45.15%	13.85	1.92	29	.06	1.85	.34	39.97–50.32
Replication 2	45.66%	8.37	2.08	15	.05	1.91	.47	41.19–50.12
Sham control	49.16%	11.34	0.30	15	.76	0.31	.08	43.11–55.20

[a] All *p*s are two-tailed. [b] *z*s are given for Stouffer *z* computations. [c] The effect size is derived from $r = \sqrt{\frac{t^2}{t^2 + df}}$.

Secondary Analyses

Linear correlation coefficients (Pearson *r*s) were calculated in order to determine the interrelationships among the magnitude of the remote-staring detection effect, SAD scoring, and MBTI extraversion-introversion (E/I) scoring. To study the relationship between remote-staring detection and SAD, Pearson *r*s were computed for the percent electrodermal activity occurring during the staring periods (as in the primary analyses) versus the SAD scores (expressed as a percentage of the highest possible SAD score) for Replication 1, for Replication 2, and for the sham control sessions. Summary statistics are provided in Table 2. For Replication 1, the magnitude of the remote-staring detection

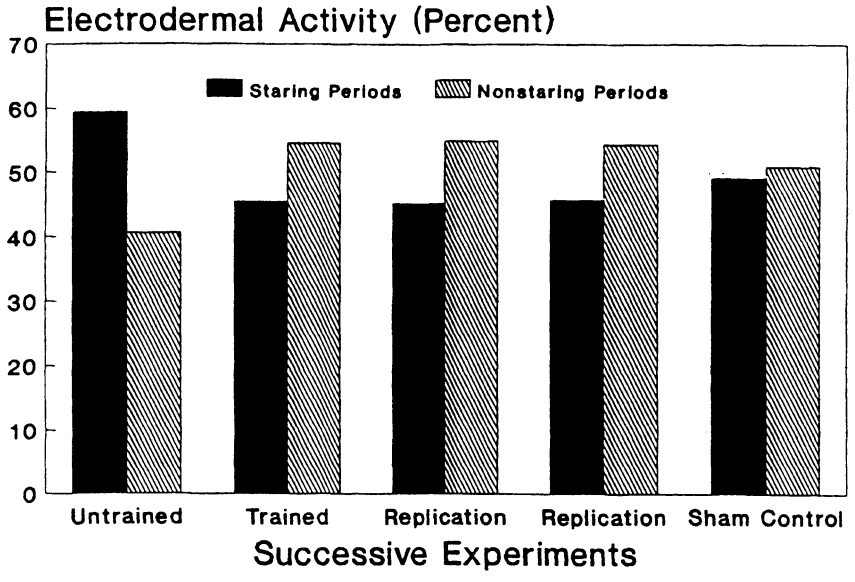

Figure 1. Percent spontaneous electrodermal activity during staring and nonstaring periods for the four experimental series and for the sham control series.

effect (the degree of "calming" during the staring periods) was significantly ad positively correlated with degree of social avoidance and distress; a similar trend was found for Replication 2. For the sham control sessions, on the other hand, this same correlation was small, negative, and nonsignificant.

TABLE 2
LINEAR CORRELATIONS BETWEEN STARING PERIOD EDA (PERCENT) AND SOCIAL AVOIDANCE AND DISTRESS (SAD) SCORE

Series	*r*	*df*	*p*[a]
Replication 1	.36	28	.05
Replication 2	.43	14	.09
Sham control	−.12	14	.66

[a] All *p*s are two-tailed.

To study the remote staring detection as related to introversion, Pearson *r*s were computed for the percent electrodermal activity during the staring periods versus the continuous score for the MBTI introversion

scale, for Replication 2 and for the sham control sessions. (The MBTI was not administered for Replication 1.) Summary statistics appear in Table 3. For comparative purposes, similar analyses are presented for our previous two series with untrained and trained starees (in which the MBTI, but not the SAD, had been administered). For Replication 2, there was a strong, positive, and highly significant correlation between the magnitude of the remote staring detection effect and the staree's degree of MBTI introversion. No such correlation occurred for the sham control segment of the experiment.

TABLE 3
LINEAR CORRELATIONS BETWEEN STARING PERIOD EDA (PERCENT) AND MBTI EXTRAVERSION/INTROVERSION (E/I) SCORE

Series	*r*	*df*	*p*[a]
Replication 2	.68	14	.0037
Sham control	.16	14	.55
Untrained	.12	14	.66
Trained	.07	14	.80

[a] All *p*s are two-tailed.

The relationships between remote staring detection and SAD scoring are presented graphically in Figure 2, and the relationships between remote staring detection and MBTI introversion scoring are presented graphically in Figure 3. In these figures, the ordinate indicates the percentage of total spontaneous electrodermal activity that occurred during the staring periods; increasing departures below the 50% chance level indicate increasing remote calming effects. The abscissas indicate, respectively, increasing degrees of social avoidance/distress/anxiety and increasing tendencies toward MBTI introversion.

For Replication 2, scores were available for both the SAD assessment and the MBTI introversion assessment, and these two instruments could be intercorrelated. The Pearson *r* for SAD versus introversion was .53 which, with 14 *df*, was associated with a two-tailed $p = .035$. The direction of the correlation was, of course, for high social avoidance and distress to be positively correlated with introversion.

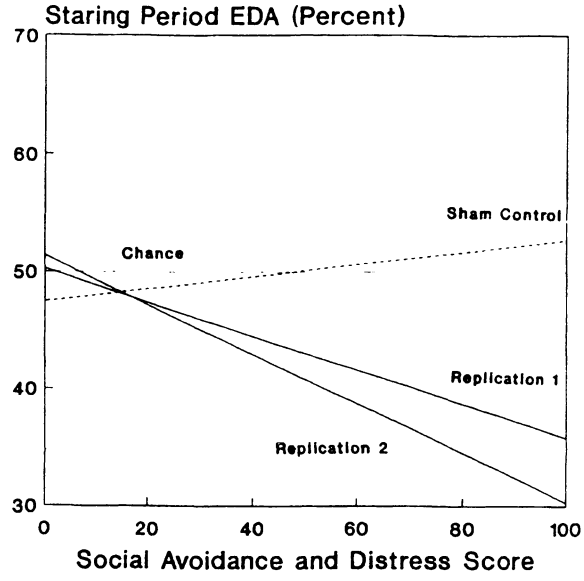

Figure 2. Linear regressions between remote staring detection (calming) and SAD scoring.

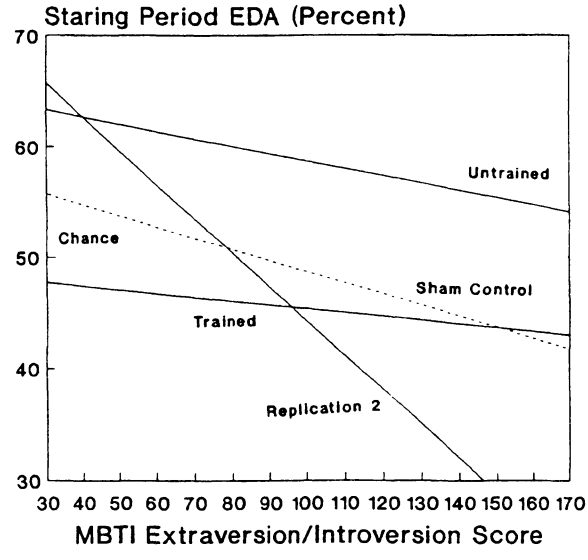

Figure 3. Linear regressions between remote staring detection (calming) and MBTI introversion scoring.

Discussion

Four separate experiments have now been carried out to determine whether persons are able to discriminate periods in which they are watched remotely by someone beyond the range of their conventional senses from periods in which such remote watching is not taking place. We reasoned that measurements of relatively unconscious autonomic nervous system activity might provide unusually sensitive indications of successful discrimination. Evidence for autonomic discrimination of remote watching or staring was indeed obtained in all four studies. This evidence reached statistical significance (as adjudged by conservative, two-tailed p values) in two earlier studies (Braud, Shafer, & Andrews, 1993) and in Replication 2 of the present paper, and very closely approached significance ($p = .06$, two-tailed) in Replication 1 of the present paper. The effect sizes (see Rosenthal, 1984, 1985) were all relatively large, ranging from .34 to .57. Inspection of Table 1 provides convincing evidence that autonomic staring detection occurred and was replicated in these studies. The absence of a similar effect in the special sham control trials provides another indication that the effect obtained in the real trials was not artifactual.

In the very first experiment (conducted with an untrained starer and untrained starees), remote staring (remote attention) was associated with autonomic activation. In the remaining three experiments, a remote autonomic calming effect was observed. We suggest the following interpretation for these different effects. Although we attempted to equate staring conditions as closely as possible in all experiments, different psychological conditions did nonetheless occur. In the very first experiment, the starer was uneasy and somewhat nervous about the prospect of staring at another person (via the closed-circuit television system) and felt she was "intruding" upon the starees. It is likely that the starer's anxiety (and therefore her heightened sympathetic nervous system activation) may have flavored her attempts to "purely attend" to the starees, and this increased activation may have been communicated to the starees. Prior to the second experiment, both starer and starees had undergone intensive connectedness training and everyone felt very comfortable and relaxed about staring and about "merging" with one another. The starer, in fact, reported feeling much more relaxed, positive, and nonanxious about her sessions in the second experiment, and it is likely that the starer's attention was flavored by these relaxed and comfortable feelings (and their associated sympathetic nervous system deactivation); and these feelings could have been communicated to the

starees in the second experiment. Additionally, the starees themselves were comfortable and relaxed about staring and merging, as a result of their own connectedness training. The starer's relaxed and comfortable state could have carried over into Replication 2 (in which she was again the starer) and could have added a relaxed character to her remote attention, even though she was now working with new starees who had not been trained. It is also likely that the starer communicated some of her relaxed attitude, characteristics, and expectations to the three starers of Replication 1 during her training of the latter and her discussions of earlier results with these starers, and that these starer characteristics were then communicated to the Replication 1 starees. Such interpretations could be tested in analytical studies in which the attitudes and conditions of starers are deliberately manipulated.

Although the apparently discrepant findings of the first versus the remaining three studies make sense in terms of the foregoing interpretation, a conservative strategy can be used in pooling the four results: the sign of the t, z, and r scores can be reversed for the result that is inconsistent with the bulk of the results, according to a recommendation by Rosenthal (1984, p. 95). This is the reason for the minus signs in the first row of Table 1.

One of the rationales for conducting these studies in the first place was not only to study staring detection using a new (and hopefully more sensitive) methodology, but also to study a pure attention component that may have been present in all of our prior biological psychokinesis experiments, along with the specific, directional, intentional aims of those experiments. The lesson of the present series of studies is that it is difficult to isolate pure attention, that the latter is easily adulterated by other starer feelings, and that the quality of the starer's attention is important in determining the nature of the experimental outcome.

The significant correlations that obtained between the remote staring detection effect and the two personality variables of SAD and MBTI introversion in the real experiments but not in the sham control segment provide additional evidence for the reality of the remote staring detection effect and also relate the magnitude of the effect to certain psychological variables. The interpretation of these relationships is still unclear; however, certain preliminary suggestions can be offered. Because of the high, significant correlation between SAD and introversion ($r = +.53$), it may be the case that one of these is functioning as a moderator variable in the interaction of the other with the magnitude of the remote-staring detection effect.

Inspection of Figure 2 reveals that as social avoidance, distress, and anxiety increase from zero to high values, the remote calming effect increases from zero (i.e., 50% MCE) to high values. We can offer the following speculative interpretations of the greater susceptibility of those starees with high social avoidance, distress, or anxiety to the remote calming effect.

1. Persons with greater social avoidance/distress/anxiety may be more sensitive to social interactions (in a vigilant way), even when those interactions are psi-mediated. Therefore, high SAD starees may have been more likely to detect the experiment's remote-staring procedure and to have responded appropriately. (Such a suggestion is consistent with an earlier empirical finding by Watson and Friend [1969] that persons with high SAD scores tended to score more highly on an "audience sensitivity index" [Paivio, 1965] than did persons with low SAD scores.)

2. Persons with greater social avoidance/distress/anxiety are ordinarily more isolated and fearful and therefore more "needy" of social interactions. Their normal need to "connect" socially with others is ordinarily denied. Perhaps their greater need for social interaction provided greater motivation for the efficacy of the remote-staring detection effect. This finding would parallel an earlier finding (Braud & Schlitz, 1983) of a greater remote mental influence effect in persons with greater "need" to be influenced. Stated somewhat differently, perhaps the remote-staring procedure of the present experiment provided a less threatening opportunity for social responsivity than is normally the case, and those persons with greater need took greater advantage of such an opportunity.

3. Persons with greater social avoidance/distress/anxiety may simply be more comfortable (than those with lower SAD scores) and at ease working alone and could, therefore, have been more at home in the isolated staree room and less distracted than persons with lower SAD scores (who might have felt unnaturally isolated and therefore in a less than optimal state of mind).

4. Persons with greater social avoidance/distress/anxiety may be more persuadable or more conforming (having developed such a coping mode as a means of anxiety reduction) than persons with lower SAD scores and this persuasibility or conformity may extend beyond the social realm to the psi realm. (Watson & Friend, 1969, discuss this correlation of SAD with persuasibility and conformity; they do not, of course, mention the possible extension to conformity with psi influences.)

Inspection of the Replication 2 regression line in Figure 3 reveals that MBTI introverts tend to exhibit remote calming, whereas MBTI extraverts tend to show a reversal of this effect (i.e., they evidenced remote activation during staring periods). As in the case of SAD scoring, we can offer some speculative interpretations of this relationship.

1. If becoming calmer is the appropriate response to remote staring under the conditions of the experiment, the more appropriate reaction of introverts may simply be due to their greater ease and comfort under the specific test conditions of the experiment (sitting alone in a room, essentially doing nothing other than "being with themselves" for about 30 min, compared with extraverts (who might be less comfortable, more restless, more distracted, and so forth, and whose less than optimal psychological state may reverse the direction of the psi effect).

2. There are empirical indications that introverts evidence greater sympathetic autonomic arousal than extraverts (Coles, Gale, & Kline, 1971; Geen, 1984; Sadler, Mefferd, & Houck, 1971), that introverts are more excitable or arousable than extraverts in response to given levels of stimulation (Geen, 1984), and that the optimal level of stimulation needed to produce a preferred level of physiological arousal may be lower for introverts than for extraverts (Eysenck, 1967). If it is hypothesized that directing remote attention or psi attention toward a person has a balancing or homeostasis-enhancing influence, then perhaps such a balancing influence would be in the direction of calming for introverts (who are naturally more "excitatory" and may be overaroused ordinarily), but in the direction of activation for extraverts (who are naturally more "inhibitory" and may be underaroused ordinarily). This hypothesis is not unrelated to LeShan's (1966) suggestion that a single moment of special attention in which one feels "at one" with another may be sufficient to trigger optimal self-healing or self-balancing events within that other.

3. To the degree that introversion is correlated with social avoidance/distress/anxiety, the various interpretations offered above in connection with SAD scoring would also be applicable to introversion.

Several of these postulated processes, along with still others, could have interacted to yield the obtained experimental outcomes. Further research would, hopefully, clarify these interrelationships.

In Replication 2, the staree's conscious attention was directed to a personally engaging task (completing personality assessments) during the experimental session. Nonetheless, the staree's more unconscious autonomic nervous system continued to maintain a connection with, and respond appropriately to, the attention and mental processes of

another, distant person (the starer). This indicates a dissociation between the two levels of knowing/reacting, as well as the possibility of going about one's individualized activities while still remaining interconnected in an important manner with others. In this experiment, neither of these complementary processes or ways of being or knowing seemed to interfere with the other. If these sorts of physiological "coherences" can be demonstrated in the laboratory, it follows that they may also be present continuously throughout life, and may indicate that while we are all, indeed, isolated individuals, we are simultaneously interconnected members of a much more inclusive, interacting, and interdependent "long body" (see Roll, 1989).

We hope other investigators will attempt to replicate these studies. We recommend the design as one that is straightforward, has already yielded consistent positive results, and addresses a very familiar psi manifestation in a manner that is readily communicable and understandable to the experimental participants and to the public at large.

References

Braud, W., & Schlitz, M. (1983). Psychokinetic influence on electrodermal activity. *Journal of Parapsychology*, **47**, 95–119.

Braud, W., & Schlitz, M. (1989). A methodology for the objective study of transpersonal imagery. *Journal of Scientific Exploration*, **3**, 43–63.

Braud, W., Shafer, D., & Andrews, S. (1993). Reactions to an unseen gaze (remote attention): A review, with new data on autonomic staring detection. *Journal of Parapsychology*, **57**, 373–390.

Briggs, K. C., & Myers, I. B. (1957). *Myers-Briggs Type Indicator Form F.* Palo Alto, CA: Consulting Psychologists Press.

Coles, M. G. , Gale, A., & Kline, P. (1971). Personality and habituation of the orienting reaction: Tonic and response measures of electrodermal activity. *Psychophysiology*, **8**, 54–63.

Eysenck, H. J. (1967). *The biological basis of personality.* Springfield, IL: Charles C. Thomas.

Geen, R. G. (1984). Preferred stimulation levels in introverts and extraverts: Effects on arousal and performance. *Journal of Personality and Social Psychology*, **46**, 1303–1312.

LeShan, L. (1966). *The medium, the mystic, and the physicist.* New York: Viking Press.

Paivio, A. (1965). Personality and audience influence. In B. Maher (Ed.), *Progress in experimental personality research, Vol. 2.* New York: Academic Press.

PETERSON, D. M. (1978). *Through the looking glass: An investigation of the faculty of extrasensory detection of being stared at.* Unpublished thesis, University of Edinburgh, Scotland.

ROLL, W. G. (1989). Memory and the long body. In L.A. Henkel & R.E. Berger (Eds.), *Research in parapsychology 1988,* (pp. 67–72). Metuchen, NJ: Scarecrow Press.

ROSENTHAL, R. (1984). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage Publications.

ROSENTHAL, R. (1965). Designing, analyzing, interpreting, and summarizing placebo studies. In L. White, B. Tursky, & G. Schwartz (Eds.), *Placebo: Theory, research and mechanisms* (pp. 110–136). New York: Guilford Press.

SADLER, T. G., MEFFERD, R. G., & HOUCK, R. L. (1971). The interaction of extraversion and neuroticism in orienting response habituation. *Psychophysiology,* **8**, 312–318.

WATSON, D., & FRIEND, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology,* **33**, 448–457.

WILLIAMS, L. (February, 1983). Minimal cue perception of the regard of others: The feeling of being stared at. Paper presented at the 10th Annual Conference of the Southeastern Regional Parapsychological Association, West Georgia College, Carrollton, GA.

Institute of Transpersonal Psychology
744 San Antonio Road
Palo Alto, CA 94303

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[22]

EXPERIMENTER EFFECTS AND THE REMOTE DETECTION OF STARING

BY RICHARD WISEMAN AND MARILYN SCHLITZ

ABSTRACT: Each of the two authors recently attempted to replicate studies in which the "receivers" were asked to psychically detect the gaze directed at them by unseen "senders." R. W.'s studies failed to find any significant effects; M. S.'s study gave positive results. The authors then agreed to carry out the joint study described in this paper, in the hope of determining why they had originally obtained such different results. The experimental design was based on each author carrying out separate experiments, but running them in the same location, using the same equipment/procedures, and drawing participants from the same subject pool. The 32 experimental sessions were divided into two sets of randomly ordered trials. Half were "stare" trials during which the experimenter directed his/her attention toward the receiver; half were "non-stare" (control) trials during which the experimenter directed his/her attention away from the receiver. The receivers' electrodermal activity (EDA) was continuously recorded throughout each session. The EDA of R. W.'s receivers was not significantly different during stare and non-stare trials. By contrast, the EDA of M. S.'s receivers was significantly higher in stare than non-stare trials. The paper discusses the likelihood of different interpretations of this effect and urges other psi proponents and skeptics to run similar joint studies.

> . . . *the experimenter effect is the most important challenge facing modern experimental parapsychology. It may be that we will not be able to make too much progress in other areas of the field until the puzzle of the experimenter effect is solved.* (Palmer, 1986, pp. 220–221.)

The apparent detection of an unseen gaze (i.e., the feeling of being stared at, only to turn around and discover somebody looking directly at you) is a common type of ostensible paranormal experience, with between 68% and 94% of the population reporting having experienced the phenomenon at least once (Braud, Shafer, & Andrews, 1993a; Coover, 1913).

Some parapsychologists have attempted to assess whether this experience is based, at least in part, on genuine psi ability. Such studies use two

The authors would like to thank the following organizations for supporting the research described in this paper: The Perrott-Warrick Fund, Cambridge University, the Institute for Noetic Sciences, UltraMind, Ltd., the Hodgson Fund, Department of Psychology, Harvard University, and the University of Hertfordshire. We are also grateful to Matthew Smith and Emma Greening for their help in running this experiment and analyzing the data, John Palmer, Dorothy Pope, and the blind reviewers for their helpful comments and suggestions.

participants: a "sender" and a "receiver." These individuals are isolated from one another, but in such a way that the sender can see the receiver. Early experiments had the sender sitting behind the receiver (Coover, 1913; Poortman, 1959; Titchener, 1898); some later studies have used one-way mirrors (Peterson, 1978) or a closed-circuit television system (Braud, Shafer, & Andrews, 1993a, 1993b; Williams, 1983). The experimental session in this type of study is divided into two sets of randomly ordered "stare" and "non-stare" trials. During stare trials the sender directs his/her attention toward the receiver; during non-stare trials the sender directs his/her attention away from the receiver. Either during or after each trial a response is made by the receiver. In early studies, the receivers made verbal guesses as to whether they believed they had been stared at; later studies have measured receivers' electrodermal activity (EDA) throughout each trial. A number of studies have obtained statistically significant differences between responses to stare and non-stare trials and in a recent review of this work, Braud, Shafer, and Andrews (1993b) concluded:

> We hope other investigators will attempt to replicate these studies. We recommend the design as one that is straightforward, has already yielded consistent positive results, and addresses a very familiar psi manifestation in a manner that is readily communicable and understandable to the experimental participants and to the public at large. (p. 408)

Both authors of the present paper previously attempted to replicate this staring effect. The first author (R. W.) is a skeptic regarding the claims of parapsychology who wished to discover whether he could replicate the effect in his own laboratory. The second author (M. S.) is a psi proponent who has previously carried out many parapsychological studies, frequently obtaining positive findings. The staring experiments carried out by R. W. showed no evidence of psychic functioning (Wiseman & Smith, 1994; Wiseman, Smith, Freedman, Wasserman, & Hurst, 1995). M. S.'s study, on the other hand, yielded significant results (Schlitz & LaBerge, 1997).

Such "experimenter effects" are common within parapsychology and are open to several competing interpretations (see Palmer, 1989a, 1989b). For example, M. S.'s study may have contained an experimental artifact absent from R. W.'s procedure. Alternatively, M. S. may have worked with more psychically gifted participants than R. W. had, or may have been more skilled at eliciting participants' psi ability. It is also possible that M. S. and R. W. created desired results via their own psi abilities, or fraud. Little previous research has attempted to evaluate these competing hypotheses. This is unfortunate, because it is clearly

important to establish why experimenter effects occur, both in terms of assessing past psi research and attempting to replicate studies in the future. For these reasons, the authors agreed to carry out a joint study in the hope of learning why our original studies obtained such dramatically different results.

Method

Design

Our joint study required M. S. and R. W. to act as separate experimenters for two different sets of trials. The two sets of trials were carried out at the same time (early October, 1995) and in the same location (R. W.'s laboratory at the University of Hertfordshire in the U.K.). In addition, the experimenters used the same equipment, drew subjects from the same subject pool, and employed exactly the same methodological procedures. The only real difference between the trials was that one set was carried out by M. S. and the other set was run by R. W. We were curious to discover if, under these conditions, we would continue to obtain significantly different results. Each study had one independent variable with two levels—stare and non-stare. The dependent variables were the receivers' EDA during the experimental session and their responses to a "belief-in-psi" questionnaire.

Participants

Thirty-two subjects (10 males and 22 females; mean age of 25.72, age range 18 – 49) acted as receivers. Thirty of these were undergraduate psychology students studying at the University of Hertfordshire. The remaining two were the authors' colleagues. M. S. and R. W. acted in a dual capacity as both experimenter and sender.

Apparatus and Materials

Layout of room. It was clearly important to minimize the possibility of any sensory leakage between sender and receiver during the experimental sessions. For this reason the receiver was located in the University's Social Observation Laboratory while the sender was located in a small room approximately 20 meters away from the laboratory (see Figure 1).

Video equipment. A Panasonic AG-450 video camera was positioned in front of the receiver and relayed an image (via a long cable connecting the two rooms) to a 14-inch JVC color TV monitor in the sender's room.

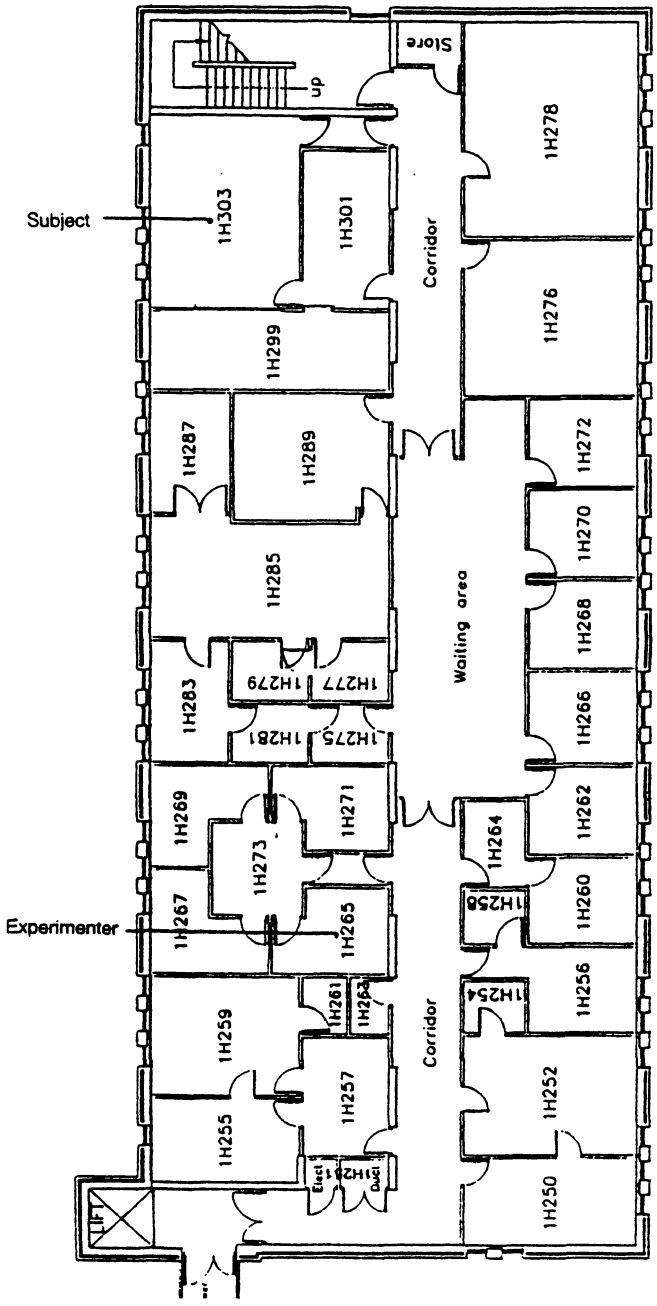

Figure 1. Locations of experimenter and subject during session.

This one-way closed circuit television system allowed the experimenter to see the subject, but not vice versa.

EDA measurement. The receivers' EDA (electrodermal activity) was recorded by the RelaxPlus system (a commercially available hardware and software package produced by UltraMind, Ltd.). This system measures skin resistance level by placing a constant current across two stainless steel electrodes and then recording the resistance encountered by that current at a rate of 10 samples per second. The system filters for possible artifacts (caused, for example, by movement) and records data to the computer's hard disk. The equipment (i.e., electrodes, input device, computer, computer monitor) was located next to the receiver throughout the experiment. The part of the program involved in storing the details of subjects and their physiological data could be accessed only via a password known only to M. S. and R. W. Data from the RelaxPlus system were then fed into a spreadsheet (Microsoft's Excel) in order to calculate the mean EDA for each 30-second trial. All statistical analyses were carried out using the Statview software package.

Belief-in-psi questionnaire. The receivers were asked three questions concerning their attitudes toward psi (see Appendix). They indicated their responses on a seven-point scale ranging from –3 to +3. A general "belief-in-psi" score was obtained by summing the receiver's responses over all three questions. Low scores on this questionnaire were taken to indicate strong belief in psi.

Trial randomization. The receivers' EDA may decline during a session for several reasons (e.g., the apparatus measuring EDA may warm up or the participants may habituate to their surroundings). This decline could lead to artifactual evidence for psi if stare trials tend to precede non-stare trials. The following randomization procedure was devised to minimize this possible artifact.

Prior to the experiment, an individual not involved in running the experiment (Matthew D. Smith) prepared a set of 32 sheets, each of which contained the order of the 32 stare or non-stare trials for one session. For 16 of these sheets the trial orders were generated in the following way: M. D. S first opened the random number table (Robson, 1983, Appendix Three), chose a number as an entry point into the table, and then threw a die twice. The numbers that came up determined how he moved from this entry point to an actual starting point. The eight consecutive numbers located in the row to the right of this starting point determined the order of the stare and non-stare trials. An even number translated into an ABBA (stare, non-stare, non-stare, stare) order while an odd number translated into a BAAB (non-stare, stare, stare, non-stare) order. The trial order for the remaining 16 sheets was determined by counterbalancing the orders of the randomized sheets just described.

Thus, a stare, non-stare, non-stare, stare on a randomized sheet became a non-stare, stare, stare, non-stare on a counterbalanced sheet. All 32 sheets were then mixed together, placed in an opaque folder, and kept in a locked drawer in R. W.'s office. M. D. S. was aware of the experimental hypotheses prior to carrying out the above randomization procedure.

Procedure

The receivers were run individually. On arriving at the laboratory, each one was met by either R. W. or M. S. Most were run by whichever of the experimenters was free to carry out the session; however, on a few occasions (e.g., when a receiver was a friend or colleague of one of the experimeters) the experimenter would be designated in advance of the trial. Thus most subjects were assigned to experimenters in an opportunistic way, rather than by one that was properly randomized (e.g., via random number tables or the output of a random number generator). The experimenter showed the subject to the receiver's room and explained the purpose of the experiment. Next, the experimenter attached electrodes to the first and third fingers of the participant's nondominant hand and made sure that the RelaxPlus system was correctly monitoring their EDA. The receivers were asked not to move their hand unnecessarily, nor to try to guess when they might be being stared at, but instead to simply remain as open as possible to any remote influence. The experimenter entered the receiver's personal data in a computerized database, initiated the recording of EDA, started a stopwatch, and left the receiver's room.

It was important that receivers were not aware of the order of the stare and non-stare trials *before* the start of the experimental session. For this reason, the list of trial orders was only selected by the experimenter only *after* he or she had left the receiver's room. The experimenter then went to R. W.'s office, retrieved the folder containing the lists of trial orders, selected any sheet he or she wanted, and proceeded to the sender's room.

Two minutes after initiating the recording of the receiver's EDA, the experimenter started to carry out the designated order of stare and non-stare trials; this order was presented to the experimenters in the form of a list. During stare trials, the experimenter quietly directed his/her attention toward the receiver; during non-stare trials the experimenter quietly directed this attention away from the receiver. Each trial lasted 30 seconds. Throughout this time the receiver completed the belief-in-psi questionnaire and then read some magazines. All of the magazines were selected to be relatively bland in content in order to minimize possible effects on the receivers' EDA.

On completion of all 32 trials, the experimenter returned to the receiver's room, thanked the participant, and told him or her that feedback of the overall results would be given within the next few weeks.

At the end of each experimental day, both experimenters copied that day's data (from their own participants as well as from the other experimenter's participants) onto their own floppy disk.

RESULTS[1]

Primary Analyses

All analyses were preplanned. A Wilcoxon signed rank test was used to compare receivers' total EDA for the 16 stare trials with their total EDA during the 16 non-stare trials.[2] Receivers run by R. W. did not differ from chance expectation (Wilcoxon $z = -.44$, $df = 15$, $p = .64$, two-tailed). In contrast, receivers run by M. S. showed a significant effect (Wilcoxon $z = -2.02$, $df = 15$, $p = .04$, two-tailed).

A "detect score" was then calculated for each subject by subtracting the total EDA during the stare trials from the total EDA for the non-stare trials. An unpaired t test revealed that the detect scores of M. S.'s subjects were not significantly different from those of R. W.'s ($df = 30$, $t = 1.39$, $p = .17$, two-tailed).

Secondary Analyses

Table 1 contains the correlation coefficients between participants' belief-in-psi questionnaire scores and their detect scores. Spearman rank correlation coefficients revealed that none of these correlations were significant. Table 1 also contains the means (and standard deviations) of

---

[1]This experiment was first reported at the 1996 Convention of the Parapsychological Association (Wiseman & Schlitz, 1996). While preparing the paper for journal publication, the authors reviewed the data and discovered an error in the way one subject's data had been transferred into the statistical package used for the analyses. For this reason the results reported here are slightly different from those reported in Wiseman and Schlitz (1996).

[2]Previous studies (e.g., Braud et al., 1993a, 1993b) have assessed their results by creating a "psi score" (the sum of EDA during stare trials divided by the sum of the total EDA) for each participant and then using a one-sample t test to determine the degree to which these scores deviate from chance expectation. This procedure obscures the question of whether an overall result is caused by a very small number of participants performing extremely well. The Wilcoxon sign rank test is more conservative than the one-sample t test because it is less influenced by the size of the deviation between participants' scores.

the questionnaire scores for R. W.'s group, M. S.'s group, and all participants.

TABLE 1
MEANS AND STANDARD DEVIATIONS FOR THE BELIEF IN PSI QUESTIONNAIRE AND CORRELATION COEFFICIENTS AND *p* VALUES BETWEEN SUBJECTS' QUESTIONNAIRE SCORES AND DETECT SCORES

	R.W's participants	M. S.'s participants	All participants
Mean	1.94	−.81	.56
Standard deviation (*SD*)	4.22	4.12	4.33
Correlation (*r*)	−.15	.32	.15
(Corrected for ties)			
z score	−.58	1.23	.84
p value, two-tailed	.56	.22	.39

DISCUSSION

Subjects run by R. W. did not respond differently to stare and non-stare trials. In contrast, participants run by M. S. were significantly more activated in stare than non-stare trials. These findings can be interpreted in several ways.

First, one might argue that M. S.'s significant results were caused by some type of experimental artifact. Several steps were taken to guard against this possibility. For example, neither the receivers nor the experimenters knew the order of the stare and non-stare trials before the start of the experiment; the location of the rooms minimized the possibility of any sender-to-receiver sensory leakage; and the randomization procedure ensured that the results were unlikely to be caused by progressive errors. This, coupled with the fact that one would expect any artifact to influence the results of both studies, suggests that M. S.'s significant results are unlikely to have been caused by a methodological error.

Second, one could argue that either R. W.'s or M. S.'s results were caused by receivers' cheating. For example, subjects could have discovered the order of stare and non-stare trials before the experimental session and altered their EDA accordingly. Alternatively, participants could have altered their data files so that they coincided with the order of stare and non-stare trials. Several factors mitigate against these

possibilities. First, such cheating would have been far from straightforward. For example, the selection of trial order was carried out a few moments before the start of the experimental session and it could only have been accessed by a participant who had installed some kind of covert monitoring equipment in the sender's room. Likewise, the computer could only be accessed if a participant had discovered a password which was known only to the experimenters. Also, neither R. W.'s or M. S.'s significant results are due to one exceptional participant, and one would therefore have to hypothesize that several participants successfully cheated.

Third, the results could have been caused by experimenter fraud. Although the experiment was not designed to make such fraud impossible, its design does mean that certain types of cheating would have been extremely unlikely. For example, neither experimenter could have decided to include data only from certain subjects because the full list of all subjects was known to both experimenters. However, more sophisticated forms of cheating were *theoretically* possible. For example, one experimenter could have substituted false sets of EDA values for subjects' actual values before the data were analyzed. Although possible, this would have been far from straightforward because subjects were frequently scheduled back-to-back (thus cutting to a minimum the time available for recording a false replacement session), and each experimenter made a back-up disk of all of the day's sessions at the end of each day (thus minimizing the possibility of an experimenter's substituting data after the day they had been recorded). In addition, no evidence of any cheating was uncovered during the running of the experiment or analysis of the data.

Fourth, one could argue that M. S. was working with a more "psychically gifted" population than R. W. was. This also seems unlikely because the receivers were assigned to the two experimenters in an opportunistic fashion.

Fifth, it is possible that M. S. was more skilled at eliciting subjects' psi ability than R. W. was. Interestingly, M. S.'s subjects scored higher on the "belief-in-psi" questionnaire than R. W.'s subjects did (although this difference just failed to reach significance: unpaired t value = 1.86, $df = 30$, $p = .072$, 2-tailed). Given that participants were opportunistically assigned to experimenters, this difference might be a reflection of the different ways in which R. W. and M. S. oriented receivers at the start of the experiment. It seems quite possible that the experimenters' own level of belief/disbelief in the existence of psi caused receivers to express different levels of belief/disbelief in psi and to have different expectations about the success of the forthcoming experimental session.

Videotapes of R. W.'s and M. S.'s induction procedures are currently being analyzed to identify differences in interaction and content.

Finally, it is also possible that both R. W. and M. S. used their own psi abilities to create the results they desired. This interpretation, if genuine, supports past research which suggests that successful experimenters (i.e., those who consistently obtain significant effects in psi studies) outperform unsuccessful ones on a variety of psi tasks (see Palmer, 1986, for a review of the literature supporting this notion).

In conclusion, this study reveals the value of developing collaborative relationships between skeptics and psi proponents. Both authors view this study as an initial step in the investigation of experimenter effects in psi research. Additional experiments would further aid our understanding of such effects. For example, it would be useful to carry out an experiment in which one experimenter interacted with the receiver and the other carried out the stare and non-stare trials during the experimental session. Such a study would help discover whether our initial interactions with the receiver or our behavior during the experimental session caused the results reported in this paper. We, the authors, hope to carry out such a study in the near future, and we urge other psi proponents and skeptics to run similar studies.

References

Braud, W., Shafer, D., & Andrews, S. (1993a). Reactions to an unseen gaze (remote attention): A review, with new data on autonomic staring detection. *Journal of Parapsychology*, **57**, 373–390.

Braud, W., Shafer, D., & Andrews, S. (1993b). Further studies of autonomic detection of remote staring: replications, new control procedures, and personality correlates. *Journal of Parapsychology*, **57**, 391–409.

Coover, J. E. (1913). The feeling of being stared at. *American Journal of Psychology*, **24**, 570–575.

Palmer, J. (1986). ESP research findings: the process approach. In H. L. Edge, R. L. Morris, J. Palmer, & J. H. Rush (Eds.), *Foundations of parapsychology* (pp. 184–222). London: Routledge & Kegan Paul.

Palmer, J. (1989a). Confronting the experimenter effect. *Parapsychology Review*, **20**, 1–4.

Palmer, J. (1989b). Confronting the experimenter effect. Part 2. *Parapsychology Review*, **20**(5), 1–5.

Peterson, D. M. (1978). Through the looking glass: an investigation of extrasensory detection of being stared at. M.A. Thesis, University of Edinburgh.

Poortman, J. J. (1959). The feeling of being stared at. *Journal of the Society for Psychical Research*, **40**, 4–12.

ROBSON, C. (1983). Experiment, design and statistics in psychology. London: Penguin Books.

SCHLITZ, M. J., & LABERGE, S. (1997). Covert observation increases skin conductance in subjects unaware of when they are being observed: A replication. *Journal of Parapsychology*, **61**, 185–196.

TITCHENER, E. B. (1898). The feeling of being stared at. *Science*, **8**, 895–897.

WILLIAMS, L. (1983). Minimal cue perception of the regard of others: The feeling of being stared at. Paper presented at the 10th Annual Conference of the Southeastern Regional Parapsychological Association, West Georgia College, Carrollton, GA. See *Journal of Parapsychology*, **47**, 59–60.

WISEMAN, R., & SCHLITZ, M. (1996). Experimenter effects and the remote detection of staring. *Proceedings of the Parapsychological Association 39th Annual Convention*, 149–157.

WISEMAN, R., & SMITH, M. D. (1994). A further look at the detection of unseen gaze. *Proceedings of the Parapsychological Association 37th Annual Convention*, 465–478.

WISEMAN, R., SMITH, M. D., FREEDMAN, D., WASSERMAN, T., & HURST, C. (1995). Two further experiments concerning the remote detection of an unseen gaze. *Proceedings of the Parapsychological Association 38th Annual Convention*, 480–492.

Dept. of Psychology
University of Hertfordshire
College Lane
Hatfield, Hertfordshire
England AL10 9AB
UK

Institute of Noetic Sciences
475 Gate Five Road
Suite 300
Sausalito, CA 94965

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[23]

The Efficacy of "Distant Healing": A Systematic Review of Randomized Trials

John A. Astin, PhD; Elaine Harkness, BSc; and Edzard Ernst, MD, PhD

Purpose: To conduct a systematic review of the available data on the efficacy of any form of "distant healing" (prayer, mental healing, Therapeutic Touch, or spiritual healing) as treatment for any medical condition.

Data Sources: Studies were identified by an electronic search of the MEDLINE, PsychLIT, EMBASE, CISCOM, and Cochrane Library databases from their inception to the end of 1999 and by contact with researchers in the field.

Study Selection: Studies with the following features were included: random assignment, placebo or other adequate control, publication in peer-reviewed journals, clinical (rather than experimental) investigations, and use of human participants.

Data Extraction: Two investigators independently extracted data on study design, sample size, type of intervention, type of control, direction of effect (supporting or refuting the hypothesis), and nature of the outcomes.

Data Synthesis: A total of 23 trials involving 2774 patients met the inclusion criteria and were analyzed. Heterogeneity of the studies precluded a formal meta-analysis. Of the trials, 5 examined prayer as the distant healing intervention, 11 assessed noncontact Therapeutic Touch, and 7 examined other forms of distant healing. Of the 23 studies, 13 (57%) yielded statistically significant treatment effects, 9 showed no effect over control interventions, and 1 showed a negative effect.

Conclusions: The methodologic limitations of several studies make it difficult to draw definitive conclusions about the efficacy of distant healing. However, given that approximately 57% of trials showed a positive treatment effect, the evidence thus far merits further study.

Ann Intern Med. 2000;132:903-910.

For author affiliations and current addresses, see end of text.

The widespread use of complementary and alternative medicine (CAM), commonly defined as therapies that are "neither taught widely in U.S. medical schools nor generally available in U.S. hospitals" (1), is now well documented. Results of several national surveys in the United States and elsewhere suggest that up to 40% of the adult population has in the preceding year used some form of CAM to treat health-related problems (1–5). In part because of the increasing use of CAM by the public, there has been a greater sense of urgency and motivation on the part of the scientific community to study the safety and efficacy of these therapies.

A belief in the role of mental and spiritual factors in health is an important predictor of CAM use (2). In a recent study of CAM in the United States (1), 7% of persons surveyed reported having tried some form of "spiritual healing." This was the fifth most frequently used treatment among all CAM therapies assessed. In the same study, 35% of persons surveyed reported that they had used prayer to address their health-related problems. A national survey conducted in the United States in 1996 found that 82% of Americans believed in the healing power of prayer and 64% felt that physicians should pray with patients who request it (6). Although not without its critics (7), a growing body of evidence suggests an association between religious involvement and spirituality and positive health outcomes (8–11).

Spiritual healing is a broad classification of approaches involving "the intentional influence of one or more persons upon another living system without utilizing known physical means of intervention" (12). Following the example of Sicher and colleagues (13), we use the term *distant healing* in our review. Although it does not necessarily imply any particular belief in or referral to a deity or higher power, distant (or distance) healing encompasses spiritual healing, prayer, and their various derivatives and has been defined as "a conscious, dedicated act of mentation attempting to benefit another person's physical or emotional well being at a distance" (13). As we define it here, distant healing includes strategies that purport to heal through some exchange or channeling of supraphysical energy. Such approaches include Therapeutic Touch, Reiki healing, and external qigong. Although they do not necessitate actual physical contact, these

healing techniques usually involve close physical proximity between practitioner and patient. Distant healing also includes approaches commonly referred to as "prayer." Prayer, whether directed toward health-related matters or other areas of life, includes several variants: intercessory prayer (asking God, the universe, or some higher power to intervene on behalf of an individual or patient); supplication, in which one asks for a particular outcome; and nondirected prayer, in which one does not request any specific outcome (for example, "Thy will be done . . .").

All forms of distant healing are highly controversial. Despite several positive reviews examining the research on these techniques (12–14), there continue to be conflicting claims in the literature regarding their clinical efficacy (7, 15, 16). In the absence of any plausible mechanism, skeptics are convinced that the benefits being reported are due to placebo effects at best or fraud at worst. Notwithstanding this ongoing controversy, distant healing techniques are increasing in popularity. For example, in the United Kingdom today, there are more distant healers (about 14 000) than there are therapists from any other branch of CAM. This level of popularity makes examination of the available evidence relevant. The objective of our systematic review was to summarize all available randomized clinical trials testing the efficacy of all forms of distant healing as a treatment for any medical condition.

Methods

A comprehensive literature search was conducted to identify studies of distant healing (spiritual healing, mental healing, faith healing, prayer, Therapeutic Touch, Reiki, distant healing, psychic healing, and external qigong). The MEDLINE, PsychLIT, EMBASE, CISCOM, and Cochrane Library databases were searched from their inception to the end of 1999. The search terms used were the above-named forms of treatment plus *clinical trials, controlled clinical trials,* and *randomized controlled trials.* In addition, we contacted leading researchers in the fields of distant and spiritual healing to further identify studies. We also searched our own files and the reference sections of articles on distant healing that we identified. Numerous studies have been carried out in these areas—for example, in a review of spiritual healing, Benor (12) identified 130 controlled investigations, and Rosa and colleagues (15) identified 74 "quantitative studies" of Therapeutic Touch. However, we included only studies that met the following criteria: 1) random assignment of study participants; 2) placebo, sham, or otherwise "patient-blindable" or adequate control interventions; 3) publication in peer-reviewed journals (excluding published abstracts, theses, and unpublished articles); 4) clinical (rather than experimental) investigations; and 5) study of humans with any medical condition.

We did not apply restrictions on the language of publication. The methodologic quality of studies was assessed by using the criteria outlined by Jadad and colleagues (17). In addition, we examined the extent to which studies were adequately powered, randomization was successful (that is, it resulted in homogenous study groups), baseline differences were statistically controlled for, and patients were lost to follow-up. Other predefined assessment criteria were study design, sample size, type of intervention, type of control, direction of effect (supporting or refuting the hypothesis), and type of result. Extracted data were entered into a custom-made spreadsheet. Differences between two independent assessors were settled by consensus. A meta-analytic approach was considered but was abandoned when the heterogeneity of the trials became apparent. Nevertheless, effect sizes averaged across each category of distant healing were included in an effort to provide some quantitative measure of the magnitude of clinical effects. Effect sizes were calculated by using Cohen's d (18), weighted for sample size. The Hedges correction was applied to all effect sizes (19). In studies that reported multiple outcomes, a single outcome was chosen to calculate effect size if 1) a significant change after treatment was shown for that outcome or 2) that outcome was the primary outcome measure in studies that found several or no significant treatment effects. In the few cases in which the authors did not provide sufficient information with which to calculate Cohen's d, the study was not included in the overall effect size.

The funding sources were not involved in the design of the study and had no role in the collection, analysis, or interpretation of the data or in the decision to submit the manuscript for publication.

Data Synthesis

Using our search methods, we found more than 100 clinical trials of distant healing. The principal reasons for excluding trials from our review were lack of randomization, no adequate placebo condition, use of nonhuman experimental subjects or nonclinical populations, and not being published in peer-reviewed journals. Twenty-three studies met our inclusion criteria (13, 20–41). These trials included 2774 patients, of whom 1295 received the experimental interventions being tested. Methodologic details and results of these trials are summarized in **Tables 1** to **3.**

The studies are categorized as three types: prayer, Therapeutic Touch, and other distant healing. How-

Table 1. Randomized, Placebo-Controlled Trials of Prayer

Author, Year (Reference)	Design	Sample Size	Experimental Intervention	Control Intervention*	Result	Comments	Jadad Score
Joyce and Welldon, 1965 (20)	Double-blind; 2 parallel groups	48 patients with psychological or rheumatic disease	Prayer in Christian or Quaker tradition; patients received 15 hours of daily prayer for 6 months	Usual care	No significant differences in clinical or attitude state	Inclusion and exclusion criteria not stated; heterogeneous patient groups; results of only 16 pairs available	5
Collipp, 1969 (21)	Triple-blind; 2 parallel groups	18 children with leukemia	Daily prayer for 15 months	Usual care	Higher death rate in control group, but difference was not significant ($P = 0.1$)	Heterogeneity of groups makes findings inconclusive, inclusion criteria not stated	4
Byrd, 1988 (23)	Double-blind; 2 parallel groups	393 coronary care patients	Prayer in Christian tradition; 3 to 7 intercessors per patient until patient was released from hospital	Usual care	Treatment group required less ventilatory support and treatment with antibiotics or diuretics	Outcomes combined into "severity score" to handle multiple comparisons; score was lower in treatment group	5
Walker et al., 1997 (24)	Double-blind; 2 parallel groups	40 patients receiving alcohol abuse treatment	Prayer for 6 months	Usual care	No treatment effect on alcohol consumption	Insufficiently powered	4
Harris et al., 1999 (39)	Double-blind; 2 parallel groups	990 coronary care patients	Remote intercessory prayer in Christian tradition for 28 days	Usual care	Significant treatment effects for summed and weighted coronary care unit score; no differences in length of hospital stay	No differences were observed when the summed scoring system developed in Byrd's study (23) was used; unclear whether baseline differences were adequately controlled for	5

* A placebo was unnecessary because patients were unaware of whether prayers were made on their behalf

ever, these classifications are not mutually exclusive. For example, the study of distant healing by Sicher and colleagues (13) included 40 healers, some of whom would describe what they did as prayer, and the study by Miller (22) described the intervention as both prayer and remote mental healing.

Prayer

Of studies that met our inclusion criteria, five specifically examined prayer as the distant healing intervention (**Table 1**). In all five studies, the intervention involved some version of intercessory prayer, in which a group of persons was instructed to pray for the patients (there was no way to control for whether patients prayed for themselves during the study). Qualifications for being an intercessor varied from study to study. For example, in the trial by Byrd (23), intercessors were required to have an "active Christian life, daily devotional prayer, and active Christian fellowship with a local church." In the study by Harris and colleagues (39), those praying were not required to have any particular denominational affiliation, but they needed to agree with the statement "I believe in God. I believe that He is personal and is concerned with individual lives. I further believe that He is responsive to prayers for healing made on behalf of the sick."

In each of these studies, the intercessors did not have any physical or face-to-face contact with the persons for whom they were praying. Instructions on how the intercessors should pray were fairly open-ended in most instances. For example, in the trial by Harris and colleagues (39), intercessors were asked to pray for a "speedy recovery with no complications and anything else that seemed appropriate to them" (39).

Two trials showed a significant treatment effect on at least one outcome in patients being prayed for (23, 39), and three showed no effect (20, 21, 24) (**Table 1**). The average effect size, computed for four of these studies, was 0.25 ($P = 0.009$).

Therapeutic Touch

Eleven trials examined the healing technique known as "noncontact Therapeutic Touch" (**Table 2**). A criterion for inclusion in our review was that the Therapeutic Touch intervention be compared to an adequate placebo, consisting of a mock or mimic Therapeutic Touch condition or a design in which patients could not physically observe whether a Therapeutic Touch practitioner was working on them. Of the 11 trials, 7 showed a positive treatment effect on at least one outcome (25, 27, 28, 30, 33, 34, 41), 3 showed no effect (26, 29, 31), and 1 showed a negative treatment effect (the controls healed significantly faster) (32) (**Table 2**). The average effect size, computed for 10 of the studies, was 0.63 ($P = 0.003$).

Table 2. Randomized, Placebo-Controlled Trials of Therapeutic Touch

Author, Year (Reference)	Design	Sample Size	Experimental Intervention	Control Intervention	Result	Comments	Jadad Score
Quinn, 1984 (25)	Double-blind	60 patients in cardiovascular unit	Noncontact Therapeutic Touch for 5 minutes	Simulated or mock Therapeutic Touch	17% decrease in posttest anxiety scores in treatment group		2
Keller and Bzdek, 1986 (27)	Single-blind, 2 parallel groups	60 patients with tension headache	Noncontact Therapeutic Touch for 5 minutes	Mock Therapeutic Touch	Treated group showed pain reduction after trial	Treatment effects were no longer present at 4 hours of follow-up; however, when participants who used intervening therapy were removed from analysis, 4-hour changes became significant	3
Quinn, 1988 (26)	Single-blind; 3 parallel groups	153 patients awaiting open-heart surgery	Noncontact Therapeutic Touch for 5 minutes	Mock Therapeutic Touch; no treatment	No significant treatment effects	Negative findings suggest importance of eye and face contact	2
Meehan, 1992 (28)	Single-blind; 3 parallel groups	108 postoperative patients	Noncontact Therapeutic Touch for 5 minutes	Mock Therapeutic Touch, usual care (analgesic drugs)	Nonsignificant reductions in postoperative pain ($P < 0.06$), treatment group showed reduced need for analgesic medication	Used conservative "intention-to-treat" analyses	3
Simington and Laing, 1993 (29)	Double-blind; 3 parallel groups	105 institutionalized elderly patients	Noncontact Therapeutic Touch with back rub for 3 minutes	Mock therapeutic touch with back rub; back rub alone	Lower levels of posttest anxiety observed in treatment group compared with back rub only	No differences between therapeutic touch and mock therapy; no pretest given	2
Wirth et al., 1993 (30)	Double-blind	24 participants with experimentally induced puncture wounds	Noncontact Therapeutic Touch (healer behind one-way mirror) 5 min/d for 10 days	No treatment (placebo not necessary)	More rapid healing in treatment group		4
Wirth et al., 1996 (32)	Double-blind; 2 parallel groups	38 participants with experimentally induced puncture wounds	Noncontact Therapeutic Touch (healer behind one-way mirror), 5 min/d for 10 days	No treatment (placebo not necessary)	No treatment effect in terms of healing of dermal wounds	Control group healed significantly faster than treatment group	3
Gordon et al., 1998 (33)	Single-blind	31 patients with osteoarthritis of knee	Noncontact Therapeutic Touch, 1 session/wk for 6 weeks	Mock Therapeutic Touch, usual care	Treatment group showed improvements in pain, health status, and function	No change in functional disability	3
Turner et al., 1998 (34)	Single-blind; 2 parallel groups	99 burn patients	Noncontact Therapeutic Touch for 5 days; time varied from 5 to 20 minutes	Mock Therapeutic Touch	Treatment group showed reductions in pain and anxiety and had lower $CD8^+$ counts		3
Wirth et al., 1994 (31)	Double-blind crossover study	25 participants with experimentally induced puncture wounds	Noncontact Therapeutic Touch with visualization and relaxation	Visualization and relaxation without Therapeutic Touch	No treatment effect	Authors note that the number of healed wounds was insufficient to compare for analyses	4
Wirth, 1990 (41)	Double-blind	44 men with experimentally induced puncture wounds	Noncontact Therapeutic Touch (healer not visible to participants), 5 min/d for 10 days	Mock Therapeutic Touch	Treatment group showed accelerated wound healing at days 8 and 16		4

Other Distant Healing

Seven studies examined some other form of distant healing (**Table 3**). Descriptions of these interventions included "distance or distant healing" (13, 37, 38, 40), "paranormal healing" (36), "psychokinetic influence" (35), and "remote mental healing" (22). Positive treatment effects were observed in four of the trials (13, 22, 35, 37), and three showed no significant effect of the healing intervention (36, 38, 40). Effect sizes were computed for five of the studies, resulting in an average effect size of 0.38 ($P = 0.073$).

Overall Effect Size

An overall effect size was calculated for all trials in which both patient and evaluator were blinded. Along with the four studies that were previously excluded because effect sizes could not be calcu-

lated, three additional trials were excluded because it was unclear whether the evaluator was blinded to the treatment condition. For the 16 remaining trials, the average effect size was 0.40 ($P < 0.001$) across the three categories of distant healing (2139 patients). A chi-square test for homogeneity was significant ($P = 0.001$), suggesting that the effect sizes were not homogeneous. Subgroup analysis revealed that effect sizes were homogeneous within the categories of prayer and other distant healing but not within the category of Therapeutic Touch studies. In this analysis, the "fail-safe N" was 63; this value represents the number of studies with zero effect that there would have to be to make the effect size results nonsignificant. It suggests that the significant findings are less likely to be the result of a "file-drawer effect" (that is, the selective reporting and publishing of only positive results).

Methodologic Issues

Owing in part to our stringent inclusion criteria, the methodologic quality of trials was fairly high; the mean Jadad score across all studies was 3.6. No clear relation emerged between the methodologic quality of the studies and whether the results were for or against the treatment. There was a trend toward studies with higher quality scores being less likely to show a treatment effect, but this correlation was weak and not statistically significant ($R = -0.15$; $P > 0.2$).

Despite the fairly high average quality of the trials, the methodologic limitations of several studies (such as inadequate power, failure to control for baseline measures, and heterogeneity of patient groups) make it difficult to draw definitive conclusions. For example, the findings reported by Collipp (21) may have resulted from a randomization problem that produced heterogeneous patient groups (two of the eight controls had myelogenous leukemia, but no patient in the experimental group had this condition). In the study by Miller (22), the positive finding of decreased systolic blood pressure in the remote mental healing group is difficult to interpret owing to the failure to control for baseline use of blood pressure medication.

The Therapeutic Touch studies carried out by Quinn (25), Keller and Bzdek (27), Turner and

Table 3. Randomized, Placebo-Controlled Trials of Other Distant Healing Methods

Author, Year (Reference)	Design	Sample Size	Experimental Intervention	Control Intervention	Result	Comments	Jadad Score
Braud and Schlitz, 1983 (35)	Single-blind within and between participants	32 participants with high levels of autonomic arousal	Distant mental influence (intention to decrease arousal with ten 30-second sessions)	No-influence control conditions	10% reduction in galvanic skin response between control and influence sessions	No effect in participants with initially low galvanic skin response levels	3
Beutler et al., 1988 (36)	Double-blind; 3 parallel groups	120 patients with hypertension	Laying on of hands by 12 healers, 20 min/wk for 15 weeks	Healing at a distance; usual care	No treatment effect	Unclear what precisely the healers did, acute increase in diastolic blood pressure after laying on of hands	4
Wirth et al., 1993 (37)	Double-blind crossover study	21 patients with bilateral asymptomatic impacted third molar who were undergoing surgery	Distance healing (Reiki, LeShan) for 15–20 minutes 3 hours after surgery	No treatment (placebo not necessary)	Treatment group showed decrease in pain intensity and greater pain relief after surgery		4
Greyson, 1996 (38)	Double-blind	40 patients with depression	Distance healing (LeShan technique)	Usual care	No treatment effect	May have been underpowered	5
Sicher et al., 1998 (13)	Double-blind; 2 parallel groups	40 patients with AIDS	Distance healing (40 healers from different spiritual traditions; each patient treated by 10 healers)	Usual care (no placebo necessary)	Healing group had fewer new AIDS-defining illnesses, less illness severity, fewer physician visits and hospitalizations, and improved mood	Mood changes may have been due to baseline differences; no apparent statistical adjustment for multiple comparisons	5
Miller, 1982 (22)	Double-blind; 2 parallel groups	96 patients with hypertension	"Remote mental healing" in Church of Religious Science tradition	No treatment (no placebo necessary)	Decrease in systolic blood pressure in treatment group	Unclear how many participants were lost to follow-up; results given for only 4 of 8 healers; use of medication not controlled for	1
Harkness et al., (40)	Double-blind	84 patients with warts	6 weeks of distant healing ("channeling of energy") by 10 healers	No treatment (no placebo necessary)	No significant treatment effect on size or number of warts	Seems that baseline values were not controlled for in analysis	5

colleagues (34), Gordon and associates (33), and Simington and Laing (29) all used single-blind methods in which the Therapeutic Touch practitioner knew whether he or she was using the actual or mock (placebo) treatment on patients. This design may have introduced bias. For example, practitioners may have consciously or unconsciously given off nonverbal cues (such as a different posture or facial expression) (14) or silently expressed higher levels of empathy to study participants that would indicate whether the treatment was actual or mock. However, blinded observers have been unable to differentiate actual noncontact Therapeutic Touch from the mock or placebo version of this therapy (33), suggesting that in these studies, the positive findings did not result from the introduction of such biases. As suggested by Quinn (26), another potential problem with the single-blind method that she and others have used is that an experienced and skilled Therapeutic Touch practitioner may in fact continue to unconsciously "manipulate energy" in some way (that is, actually perform Therapeutic Touch), thereby producing a therapeutic effect even though his or her conscious intention is to pretend to do the procedure.

In two of the methodologically better studies that examined prayer (23) and distant healing (13), the positive findings may have resulted from a failure to use a Bonferroni correction to adjust for multiple statistical comparisons. Byrd (23), in an effort to address this problem, combined the various treatment outcomes into a "severity score" in his study and the prayer treatment group had significantly lower severity scores. Targ (Personal communication) reported that in the study by Sicher and colleagues (13), post hoc analyses in which corrections for multiple comparisons were made did not alter their results.

In studies that failed to show a significant treatment effect, weaknesses in study design (such as inadequate sample sizes) may have increased the likelihood of a type 2 error (failure to reject the null hypothesis when it is in fact false). Post hoc analyses suggested that lack of statistical power may explain the negative findings in Walker and colleagues' study of prayer (24) and Greyson's study of distant healing (38).

Finally, in Simington and Laing's (29) study of noncontact Therapeutic Touch in institutionalized elderly patients, the investigators did not collect pretest data to control for the possibility of a "testing" effect. However, without such data, it is impossible to know whether the randomization procedure actually produced homogeneous groups at baseline, which makes the post-test data difficult to interpret.

Discussion

In our systematic review of 23 randomized, controlled trials of all forms of distant healing, 13 (57%) showed a positive treatment effect, 9 showed no effect, and 1 showed a negative effect. The numbers of prayer and distant healing studies with positive and negative findings were roughly equal, whereas a somewhat larger proportion of Therapeutic Touch trials (7 of 11) showed a significant treatment effect. Results of our quantitative analysis suggest that effect sizes were small (0.25 for prayer and 0.38 for "other" distant healing) to moderate (0.63 for Therapeutic Touch). An overall statistically significant effect size of 0.40 was found across all categories of distant healing (16 trials) in which both patients and evaluators were adequately blinded.

A major limitation of our review was the heterogeneity of the trials (both in terms of treatment and outcomes), which precluded formal quantitative analyses. Furthermore, despite our restrictive inclusion criteria, we identified several methodologic limitations in the trials that made qualitative interpretation of the findings difficult. Thus, the results of our review must be interpreted with caution.

Previous reviews of distant healing techniques have also had mixed results. For example, a recent review of Therapeutic Touch for wound healing found 5 studies (all by the same author) and concluded that "results are far from impressive... [and] inconsistent overall..." (42) (The Cochrane Collaboration is currently examining the evidence for Therapeutic Touch in wound healing [43]). However, a more recent meta-analysis of 9 randomized, controlled trials of Therapeutic Touch (44) concluded that Therapeutic Touch is more effective than mock Therapeutic Touch or routine clinical touch in reducing anxiety symptoms. In addition, a meta-analytic review of 13 trials (which differed from those included in our review owing to the inclusion criteria) found an average effect size of 0.39 (45). Our findings are in basic agreement with a recent Cochrane Collaboration systematic review (46) that included results of three of the prayer trials that we reviewed (20, 21, 23) and found no clear evidence for or against the incorporation of prayer into medical practice.

As noted, the scientific investigation of such techniques as prayer, energy healing, and psychic or distant healing is controversial. One might argue that at the very least, distant healing has a powerful placebo effect that could be used to benefit certain patients in clinical practice. This would be true if we could be certain that such techniques were devoid of serious adverse events. However, O'Mathuna (16) has suggested that this may not be the case. He notes that some of the original writings of the de-

velopers of Therapeutic Touch state that patients may be harmed if they are, for example, "flooded" with too much energy. This "overdosing" of energy may manifest as irritability, restlessness, anxiety, or increased pain. O'Mathuna acknowledges, however, that these potential negative side effects of Therapeutic Touch are only speculative and have never been scientifically documented. In arguing for the importance of obtaining full consent in distant healing studies, Dossey (47) notes that some evidence in the literature indicates that distant mental influence can cause harm in nonhuman biological systems; thus, prayer and energy healing may not always be benign.

Studies of prayer also raise certain philosophical issues (48), such as why a benevolent God or deity would respond only to the prayers of or on behalf of persons in the treatment group, when many persons in the control group will probably pray for themselves and will be prayed for by friends and loved ones. Similarly, why would a compassionate God or higher power who intends the well-being of all humankind respond only to the needs of those who pray or are prayed for?

Others find the scientific scrutiny of things religious and spiritual to be misguided and even potentially blasphemous; they ask, for example, how science could ever prove or disprove the existence of things that believers take as matters of faith. Although such reservations are duly noted, we believe that there is no compelling reason why the scientific method cannot be applied to such areas as distant healing and prayer and that doing so will only further our knowledge about the potential value of these approaches in health and in life. In the words of a leading researcher in this field (48),

> No experiment can prove or disprove the existence of God, but if in fact [mental] intentions can be shown to facilitate healing at a distance, this would clearly imply that human beings are more connected to each other and more responsible to each other than previously believed. That connection could be actuated through the agency of God, consciousness, love, electrons, or a combination. The answers to such questions await further research.

Directions for Future Research in Distant Healing

As noted earlier, the studies of distant healing reviewed here have several methodologic limitations. We highlight some of the difficulties inherent in research on distant healing and offer some suggestions that might help guide future investigations into these areas.

First, as noted by both critics and proponents of distant healing, it is difficult to obtain "pure" control groups in distant healing research. For example, in prayer studies, particularly those involving very ill patients, the controls who are not being experimentally prayed for or sent healing intentions as part of the study are likely to nonetheless receive prayers and positive mental intentions from friends, loved ones, and others. We concur with Dossey (47) and others who have suggested that one solution to this seemingly unavoidable methodologic problem in such research is to carry out distant healing studies on nonhuman populations (such as animals or bacteria). The findings of controlled trials of distant healing (12) in nonhuman biological systems are provocative enough to merit further research.

Second, we agree with Targ (14), who has suggested that future studies of prayer and distant healing should more carefully measure psychological factors (such as depression, anxiety, sense of control, and self-efficacy) that are known to interact with physical health outcomes.

Third, as noted, the negative findings in many of the healing studies we reviewed may have resulted from inadequate sample sizes and insufficient statistical power. However, well-designed randomized, controlled trials of prayer and distant healing with significantly larger samples (more than 1000 patients) are in progress at several institutions (46, 49).

Fourth, in an effort to explain some of the negative findings in distant healing research, it has been suggested that blinding to assignment in randomized, controlled trials might block receptivity to "healing energy" by generating uncertainty in patients (47). Carrying out studies in nonhuman populations would, in theory, be one way to minimize this methodologic issue. Another way to test this theory would be to inform experimental and control patients that they will be receiving the distant or spiritual healing and then directly examine the extent to which patients' beliefs or "receptivity" influence study outcomes. However, ethical considerations of informed consent might make this design difficult to implement. Yet another possibility would be to design randomized, controlled trials with nonrandomized "preference arms" that would allow evaluation of the effects of randomization as opposed to choice.

Finally, it has been suggested (50–52) that previous (skeptical) beliefs of trial volunteers or investigators—the "experimenter effect"—might contribute to unsuccessful outcomes (that is, if mental intentions influence physical matter in some way, the investigators' or patients' negative beliefs about healing could directly affect study outcomes). Again, such a hypothesis could, in theory, be empirically tested by having investigators who are skeptical of and believers in spiritual healing conduct the same trials and assess whether in fact such beliefs influence outcomes (52).

Conclusions

Despite the methodologic limitations that we have noted, given that approximately 57% (13 of 23) of the randomized, placebo-controlled trials of distant healing that we reviewed showed a positive treatment effect, we concur with the summary conclusion of the Cochrane Collaboration's review of prayer studies that the evidence thus far warrants further study (46). We believe that additional studies of distant healing that address the methodologic issues outlined above are now called for to help resolve some of the discrepant findings in the literature and shed further light on the potential efficacy of these approaches.

From University of Maryland School of Medicine, Baltimore, Maryland; and University of Exeter, Exeter, United Kingdom.

Grant Support: By the National Center for Complementary and Alternative Medicine, National Institutes of Health (1P50AT0008401), The Wellcome Trust (050836/Z/970, and a charitable donation from the Maurice Laing Foundation.

Requests for Single Reprints: John A. Astin, PhD, Complementary Medicine Program, Kernan Hospital Mansion, 2200 Kernan Drive, Baltimore, MD 21207-6697; e-mail, jastin@compmed.ummc.umaryland.edu.

Requests To Purchase Bulk Reprints (minimum, 100 copies): Barbara Hudson, Reprints Coordinator; phone, 215-351-2657; e-mail, bhudson@mail.acponline.org.

Current Author Addresses: Dr. Astin: Complementary Medicine Program, Kernan Hospital Mansion, 2200 Kernan Drive, Baltimore, MD 21207-6697.
Ms. Harkness and Dr. Ernst: Department of Complementary Medicine, University of Exeter, 25 Victoria Park Road, Exeter EX2 4NT, United Kingdom.

References

1. **Eisenberg DM, Davis RB, Ettner SL, Appel S, Wilkey S, Van Rompay M, et al.** Trends in alternative medicine use in the United States, 1990-1997: results of a follow-up national survey. JAMA. 1998;280:1569-75.
2. **Astin JA.** Why patients use alternative medicine: results of a national study. JAMA. 1998;279:1548-53.
3. **Thomas KJ, Carr J, Westlake L, Williams BT.** Use of non-orthodox and conventional health care in Great Britain. BMJ. 1991;302:207-10.
4. **Millar WJ.** Use of alternative health care practitioners by Canadians. Can J Public Health. 1997;88:154-8.
5. **MacLennan AH, Wilson DH, Taylor AW.** Prevalence and cost of alternative medicine in Australia. Lancet. 1996;347:569-73.
6. **Wallis C.** Faith and healing: can prayer, faith and spirituality really improve your physical health? A growing and surprising body of scientific evidence says they can. Time. 1996;147:58.
7. **Sloan RP, Bagiella E, Powell T.** Religion, spirituality, and medicine. Lancet. 1999;353:664-7.
8. **Levin JS.** Religion and health: is there an association, is it valid, and is it causal? Soc Sci Med. 1994;38:1475-82.
9. **Kass JD, Friedman R, Leserman J, Zuttermeister PC, Benson H.** Health outcomes and a new measure of spiritual experience. Journal for the Scientific Study of Religion. 1991;30:203-11.
10. **Matthews DA, McCullough ME, Larson DB, Koenig HG, Swyers JP, Milano MG.** Religious commitment and health status: a review of the research and implications for family medicine. Arch Fam Med. 1998;7:118-24.
11. **Luskin FM, DiNucci E, Newell K.** A review of the effect of spiritual and religious factors on mortality and morbidity with a focus on cardiovascular and pulmonary disease. J Cardiopulm Rehabil. 2000;20:8-15.
12. **Benor D.** Survey of spiritual healing research. Complementary Medical Research. 1990;4:9-33.
13. **Sicher F, Targ E, Moore D 2d, Smith HS.** A randomized double-blind study of the effect of distant healing in a population with advanced AIDS. Report of a small scale study. West J Med. 1998;169:356-63.
14. **Targ E.** Evaluating distant healing: a research review. Altern Ther Health Med. 1997;3:74-8.
15. **Rosa L, Rosa E, Sarner L, Barrett S.** A close look at therapeutic touch. JAMA. 1998;279:1005-10.
16. **O'Mathuna D.** Therapeutic touch: what could be the harm? Scientific Review of Alternative Medicine. 1998;2:56-62.
17. **Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds D, Gavaghan DJ, et al.** Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996;17:1-12.
18. **Cohen J.** Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Erlbaum; 1988.
19. **Hedges LV.** Estimation of effect size from a series of independent experiments. Psychol Bull. 1982;92:490-9.
20. **Joyce CR, Welldon RM.** The objective efficacy of prayer: a double-blind clinical trial. J Chronic Dis. 1965;18:367-77.
21. **Collipp PJ.** The efficacy of prayer: a triple-blind study. Med Times. 1969;97:201-4.
22. **Miller RN.** Study on the effectiveness of remote mental healing. Med Hypotheses. 1982;8:481-90.
23. **Byrd RC.** Positive therapeutic effects of intercessory prayer in a coronary care unit population. South Med J. 1988;81:826-9.
24. **Walker SR, Tonigan JS, Miller WR, Corner S, Kahlich L.** Intercessory prayer in the treatment of alcohol abuse and dependence: a pilot investigation. Altern Ther Health Med. 1997;3:79-86.
25. **Quinn JF.** Therapeutic touch as energy exchange: testing the theory. ANS Adv Nurs Sci. 1984;6:42-9.
26. **Quinn JF.** Therapeutic touch as energy exchange: replication and extension. Nurs Sci Q. 1989;2:79-87.
27. **Keller E, Bzdek VM.** Effects of therapeutic touch on tension headache pain. Nurs Res. 1986;35:101-6.
28. **Meehan TC.** Therapeutic touch and postoperative pain: a Rogerian research study. Nurs Sci Q. 1993;6:69-78.
29. **Simington JA, Laing GP.** Effects of therapeutic touch on anxiety in the institutionalized elderly. Clin Nurs Res. 1993;2:438-50.
30. **Wirth DP, Richardson JT, Eidelman WS, O'Malley AC.** Full thickness dermal wounds treated with non-contact therapeutic touch: a replication and extension. Complement Ther Med. 1993;1:127-32.
31. **Wirth DP, Barrett MJ, Eidelman WS.** Non-contact therapeutic touch and wound re-epithelialization: an extension of previous research. Complement Ther Med. 1994;2:187-92.
32. **Wirth DP, Richardson JT, Martinez RD, Eidelman WS, Lopez ME.** Non-contact therapeutic touch intervention and full-thickness cutaneous wounds: a replication. Complement Ther Med. 1996;4:237-40.
33. **Gordon A, Merenstein JH, D'Amico F, Hudgens D.** The effects of therapeutic touch on patients with osteoarthritis of the knee. J Fam Pract. 1998;47:271-7.
34. **Turner JG, Clark AJ, Gauthier DK, Williams M.** The effect of therapeutic touch on pain and anxiety in burn patients. J Adv Nurs. 1998;28:10-20.
35. **Braud W, Schlitz M.** Psychokinetic influence on electrodermal activity. The Journal of Parapsychology. 1983;47:95-119.
36. **Beutler JJ, Attevelt JT, Schouten SA, Faber JA, Dorhout Mees EJ, Geijskes GG.** Paranormal healing and hypertension. Br Med J (Clin Res Ed). 1988;296:1491-4.
37. **Wirth DP, Brenlan DR, Levine RJ, Rodriguez CM.** The effect of complementary healing therapy on postoperative pain after surgical removal of impacted third molar teeth. Complement Ther Med. 1993;1:133-8.
38. **Greyson B.** Distance healing of patients with major depression. Journal of Scientific Exploration. 1996;10:447-65.
39. **Harris WS, Gowda M, Kolb JW, Strychacz CP, Vacek JL, Jones PG, et al.** A randomized, controlled trial of the effects of remote, intercessory prayer on outcomes in patients admitted to the coronary care unit. Arch Intern Med. 1999;159:2273-8.
40. **Harkness EF, Abbot NC, Ernst E.** A randomized clinical trial of distant healing for peripheral common warts. Am J Med. [In press]
41. **Wirth DP.** The effect of non-contact therapeutic touch on the healing rate of full thickness dermal wounds. Subtle Energies. 1990;1:1-20.
42. **Daley B.** Therapeutic Touch, nursing practice and contemporary cutaneous wound healing research. J Adv Nurs. 1997;25:1123-32.
43. **O'Mathuna D.** Therapeutic touch for wound healing. The Cochrane Database of Systematic Reviews. 1999;2.
44. **Warber SL, Gillespie BW, Klle GL, Gorenflo D, Bolling SF.** Meta-analysis of the effects of Therapeutic Touch on anxiety symptoms. Focus on Alternative and Complementary Therapies (FACT). [In press]
45. **Winstead-Fry P, Kljek J.** An integrative review and meta-analysis of therapeutic touch research. Altern Ther Health Med. 1999;5:58-67.
46. **Roberts L, Ahmed I, Hall S, Sargent C.** Intercessory prayer for the alleviation of ill health. The Cochrane Database of Systematic Reviews. 1999;2.
47. **Dossey L.** The return of prayer. Altern Ther Health Med. 1997;3:10-7.
48. **Targ E, Thomson KS.** Can prayer and intentionality be researched? Should they be? Altern Ther Health Med. 1997;3:92-6.
49. **Krucoff MW, Crater SW, Green CL, Maas AC, Seskevich JE, et al.** Randomized integrative therapies in interventional patients with unstable angina. the Monitoring and Actualization of Noetic Training (MANTRA) feasibility pilot [Abstract]. Circulation 1998;98:A1458.
50. **Brown CK.** Spiritual healing in general practice. using a quality of life questionnaire to measure outcome. Complement Ther Med. 1995;3:230-3.
51. **Wirth DP.** The significance of belief and expectancy within the spiritual healing encounter. Soc Sci Med. 1995;41:249-60.
52. **Wiseman R, Schlitz M.** Experimenter effects and the remote detection of staring. The Journal of Parapsychology. 1997;61:197-201

[24]

Give the Null Hypothesis a Chance

Reasons to Remain Doubtful about the Existence of Psi

James E. Alcock

Is there a world beyond the senses? Can we perceive future events before they occur? Is it possible to communicate with others without need of our complex sensory-perceptual apparatus that has evolved over hundreds of millions of years? Can our minds/souls/personalities leave our bodies and operate with all the knowledge and information-processing ability that is normally dependent upon the physical brain? Do our personalities survive physical death?

Experience suggests to many people that the answer to such questions is 'Yes'. Indeed, year after year, surveys show that the majority of people believe that such paranormal phenomena exist, and personal experience is one of the primary reasons for their belief. Many such experiences are emotionally powerful and bring with them meaning and existential comfort.

What accounts for these reported experiences? Do they really, some of them at least, reflect a reality beyond the materialistic world as it is now understood by science — that is, are they really 'paranormal'? Or are they the product of normal but misunderstood brain function? That is, do our brains sometimes produce or interpret experiences in such a way that they seem to be paranormal even though they are not? Parapsychologists are motivated by and large by the former interpretation, and seek scientific evidence to support that view. Mainstream science, on the other hand, takes the latter stand, usually rejecting out of hand any paranormal claims.

Whatever the explanation, given that these experiences appear to be relatively common and are often very striking, they merit study in their own right. Unfortunately, such study is rather rare. Most psychologists, eschewing paranormal and supernatural claims, have by and large ignored such experiences, while parapsychologists, on the other hand, give scant attention to normal explanations and focus instead on the paranormal possibilities. Thus, what should be of common interest to both psychologists and parapsychologists instead falls through the cracks, with one camp persuaded that the paranormal is real and the explanation

Correspondence: James E. Alcock, Dept. of Psychology, Glendon College, York University, Toronto, Canada. *Email: jalcock@yorku.ca*

for many such experiences, and the other camp rejecting the paranormal while also ignoring the experience.

As a result, parapsychologists and sceptical scientists most often speak to each other in a *dialogue aux sourds*, a dialogue of the deaf. Yet, it is always a good thing to try to build bridges in the hope of bringing intellectual protagonists together, and this special issue of the *Journal of Consciousness Studies*, which includes articles by some of the leading proponents and critics of parapsychology, may help build such a bridge. As much as they may differ in terms of their views on the paranormal, it is important to note that the contributors are 'all on the same side' in at least one important way: all share a deep respect for science and are committed to the scientific method as the appropriate approach to exploring reality. They are all seeking truth, not delusion; fact, not fiction. Arguably, the only significant differences that distinguish the proponents from the sceptics in this collection of articles are in terms of their *a priori* subjective weighings of the likelihood that psychic phenomena exist, which in turn may influence their evaluations of the adequacy of the research protocols that have been employed in parapsychological research and the quality of the data thus obtained.

Those in the scientific community who have little familiarity with parapsychology are often unaware of the wide spectrum of opinion, expertise and degree of respect for science, that exists amongst those who call themselves parapsychologists. At one end are those described in the last paragraph, of whom some have contributed to this volume. At the other are numerous writers and researchers who view science as an inadequate tool for grappling with the mysteries of the paranormal, and who base their beliefs in the paranormal solely on the kinds of experiences served up by trance mediums, putative apparitions, and so forth. Their writings are not to be found in this Special Issue, nor are the writings of those who believe that the verdict is already in, that parapsychology has long since established a sound scientific footing for paranormal phenomena and no controversy remains. (Indeed, this touches on a demarcation problem, in that scholarly, research-oriented parapsychologists reserve the label of 'parapsychologist' for themselves, and do not consider members of the general public who use this title to be parapsychologists. This important distinction is often difficult to make for those outside parapsychology.)

There is also a spectrum of opinion, expertise, and yes, of the degree of respect for science, amongst sceptics too, and again, those at the far end who only sneer dismissively at any mention of the paranormal, or those whose dogmatism shows an inability or unwillingness to be objective, are not to be found in this Special Issue.

Thus, to the sceptical reader, I stress that these parapsychological writers are in our camp, the scientific camp. They believe in science and strive to apply it. To the reader who leans towards belief in the paranormal, the sceptical writers you will find here are not motivated by any desire to drive parapsychology into the desert, but only by the desire to truly understand human experience.

That being said, I have myself long been a critic of parapsychological research, and it is only fair that I state my views 'up front'. I have yet to find any

empirical evidence that persuades me that it is likely that paranormal phenomena actually exist. Moreover, I am well aware of just how often our brains can mislead us, and can lead us to believe that we have had a paranormal experience even when no such thing has happened. Indeed, even if there is no such thing as a paranormal phenomenon, human information processing works in such a way that we are all likely from time to time to have experiences that *seem* for all the world to be paranormal. For me as a psychologist, these experiences themselves — the reports of extrasensory perception and the like — are fascinating in their own right, even if, as I presume, they are not paranormal, for they can tell us a great deal about how our brains work and about our beliefs and needs and expectations, if we are willing to listen.

I approached my own reading of the articles in this Special Issue in part with the personal desire to find out if there is any new and compelling evidence that might nudge me away from my strong scepticism about the existence of paranormal phenomena. There are for me a number of reasons to be doubtful about the existence of paranormal phenomena (I shall adumbrate some of the more important ones below), and thus, I perused each article against the backdrop of those concerns, and considered whether its conclusions supported the Psi hypothesis (that psychic, or 'psi', phenomena exist), or were they more in line with the Null hypothesis (that is, that the observed results came about naturally, and had nothing to do with psi). I advocate that the reader take a similar approach, keeping in mind not just the Psi-hypothesis, but the Null hypothesis as well.

Reasons to Remain Doubtful about the Existence of Psi

1. Lack of definition of subject matter

One of parapsychology's most vexing problems has to do with the very definition of its subject matter. What is it that is being studied, and how are the phenomena under study themselves defined? Are ghosts, levitation and trance channellers part of the accepted range of subject matter? Or is the subject matter at this time restricted to subtle mind-induced influences at the micro level? If mainstream science is challenged to consider seriously the claims of parapsychology, just what claims are we talking about — ghost sightings, or small but statistically significant changes in a distribution of outcomes of a random event generator in a laboratory experiment? Parapsychological opinion as to its proper subject matter varies widely. Consider, for example, the views of the following organizations:

The Parapsychological Association (PA) defines itself as 'the international professional organization of scientists and scholars engaged in the study of "psi" (or "psychic") experiences, such as telepathy, clairvoyance, remote viewing, psychokinesis, psychic healing, and precognition . . .', and its webpage states:

> The diversity found within PA membership also leads to many different 'schools of thought' regarding the phenomena studied — ranging from those who suspect that psi will eventually turn out to be an artifact of no major significance, to those who

> believe it will be accounted for through new developments in physics or biology, to those who argue that psi phenomena suggest a basis for spiritual beliefs.
>
> (*www.parapsych.org*)

In Britain, the venerable Society for Psychical Research states that:

> The principal areas of study of psychical research concern exchanges between minds, or between minds and the environment, which are not dealt with by current orthodox science. This is a large area, incorporating such topics as extrasensory perception (telepathy, clairvoyance, precognition and retrocognition), psychokinesis (paranormal effects on physical objects, including poltergeist phenomena), near-death and out-of-the-body experiences, apparitions, hauntings, hypnotic regression and paranormal healing. One of the Society's aims has been to examine the question of whether we survive bodily death, by evaluating the evidence provided by mediumship, apparitions of the dead and reincarnation studies.
>
> (*www.spr.ac.uk/about.html*)

The American Society for Psychical Research (*www.aspr.com/topics.htm*), which describes itself as the oldest psychical research institute in the United States, lists as its subject matter an extensive range of topics, from extrasensory perception to psychic healing to trance channellers and survival after death to dowsing and poltergeists.

The point is that there is a great variety of opinion as to what constitutes the essential and appropriate subject matter of parapsychology. Some parapsychologists want to adhere to the rules of evidence as they exist in modern science, while others rely on anecdotal accounts of wondrous events — such as the supposed levitation of the medium Daniel Douglas Home during a seance in 1852 — as the best evidence that psi phenomena are real. Recently, some prominent members of the Society for Psychical Research have become very interested in the study of spirit mediums once again, based on sittings in Scole, England, and view this evidence as strongly suggestive of communication with a spirit world (Keen, Ellison and Fontana, 1999). Yet, such research would probably be considered quaint and unscientific by more laboratory-oriented parapsychologists such as those whose articles appear herein, and such research, despite the importance it is given in some parapsychological circles, is not even mentioned by the parapsychologists who have contributed to this Special Issue.

This all reflects the fact that to the extent that parapsychology constitutes a 'field' of research, it is a field without a core knowledge base, a core set of constructs, a core set of methodologies, and a core set of accepted and demonstrable phenomena that all parapsychologists accept. Moreover, I consider it doubtful that parapsychologists could agree amongst themselves as to just what experiments or demonstrations in the literature constitute the best case for psi. This immediately distinguishes parapsychology from any other scientific research field, where there is always a common core of knowledge as well as key demonstrations that can reliably be produced and taught, even while there may be controversy about various concepts and research findings at the frontiers of the field.

2. Definition of constructs

Quite apart from differences of viewpoint in what constitutes the range of appropriate subject matter, a much more important definitional problem arises in terms of defining and measuring specific psi phenomena. The problem arises primarily because psi phenomena are defined, not in terms of what they are, but only in terms of what they are not. Telepathy is the simultaneous sharing or transfer of information between two brains *in the absence of any 'normal' mechanism that could account for it*; precognition involves seeing future events *in a manner that cannot be accounted for by any means understood by contemporary science*, and so on. Telepathy is not telepathy if sender and receiver communicate by 'silent' dog whistles that one of them is able to hear, or if they have some sort of secret code that allows them to communicate without the knowledge of the researcher. Psychokinesis is not psychokinesis if the psychic causes an object to move by hidden, although normal, means. Indeed, parapsychology is the only realm of objective inquiry in which the phenomena are all *negatively* defined, defined in terms of ruling out normal explanations. Of course, ruling out all normal explanations is not an easy task. We may not be aware of all possible normal explanations, or we may be deceived by our subjects, or we may deceive ourselves.

If all normal explanations actually could be ruled out, just what is it that is at play? What is *psi*? Unfortunately, it is just a label. It has no substantive definition that goes beyond saying that all normal explanations have apparently been eliminated. Of course, parapsychologists generally presume that it has something to do with some ability of the mind to transcend the laws of nature as we know them, but all that is so vague as to be unhelpful in any scientific exploration. Some parapsychologists, recognizing the problem of trying to provide a positive rather than a negative definition of psi, choose to sidestep the issue and instead focus on 'anomalies'. Psi effects are thus thought of as anomalous findings that apparently should not occur if the current scientific worldview is accurate. These are not just any such anomalies, of course. They are anomalies that involve, in one way or another, the mind.

Anomalistic observations that do not fit with accepted theory are vital to scientific progress, for they force us to modify our theories and to gather additional data until they can be understood and accommodated into a revised theory. For example, to AIDS researchers it is quite anomalous that some Nairobi prostitutes show an inherent resistance to HIV infection, but only as long as they continue to have exposure to multiple partners. This is an important anomaly — it does not make immediate sense in terms of what is known about this illness, but coming to understand it will undoubtedly lead to a much better understanding of HIV in general. Elsewhere in science, anomalies sometimes lead to such fundamental changes in theory that philosophers of science speak in terms of a paradigm shift. The precession of Mercury in its orbit behind the sun was anomalous; for it did not fit with Newton's theory of gravity and the derivative understanding of the movement of planets. Scientists a century ago went so far as to speculate that Mercury's orbit behind the sun was actually disrupted by the gravitational field of an unseen planet (they called it Vulcan) on the far side of the sun. However,

Einstein's general theory of relativity was able to account for the perihelion shift of Mercury, resolving the anomaly and thereby helping to usher in a new scientific worldview.

Yet, when parapsychologists seek to establish their subject matter in terms of anomalies, there is something quite different going on compared to either of the examples above. In mainstream science, one does not deliberately seek anomalies; they present themselves. They are unexpected and unpredicted by current theory, that is why, after all, they are called anomalies. However, no psi anomaly has ever presented itself in the course of research in mainstream science. Consider the particularly delicate experiments in subatomic physics, which might be ideal for the manifestation of putative psi forces, given that they involve very tiny amounts of matter and energy, highly precise measurements and very highly motivated researchers with, at least at times, varying expectations. We do not read research reports that suggest that the outcomes of such experiments seem to depend on who was operating the linear accelerator at the time, and that a particular effect is found only when certain researchers are present and not otherwise, reflecting perhaps a researcher's 'psychic' influence. In the course of doing normal science, anomalies suggestive of psi just do not pop up. Rather, parapsychologists, in their work, deliberately try to generate them; they are the goal of much parapsychological research and are only labelled as anomalous by the rather circular route of deeming them to be impossible if current science is accurate and complete.

Parapsychologists need to be able to provide a positive definition of psi, to tell us how to identify psi 'anomalies' in ways other than exclusion, and to tell us how to rule out psi, how to know when it is absent. This problem is as great now as it has ever been, and no progress has been made in overcoming it across more than a century of empirical parapsychological research. Because of its negative definition, we are left with no idea as to when psi might occur, and more importantly to the scientist, as to when it will *not* occur. There is no way, we are told, that psi can be blocked or attenuated by the researcher, and thus we cannot compare conditions where psi could not occur to those where, were it to exist, it could be observed. Moreover, because it is claimed that psi influences can occur without any attenuation as a function of distance, and can occur backwards and forwards in time, it becomes impossible ever to truly 'control' the conditions of an experiment.

3. Failure to achieve replication

If parapsychologists cannot provide a positive definition of psi, then at least one would hope that they could provide a reliable, replicable, demonstration of the subject of their study, be it an 'anomaly' or whatever. Mainstream science accords a high value to replicability, for it is perhaps the best safeguard against being taken in by results produced by error, self-delusion or fraud. Yet replicability itself is a somewhat complex concept. Simply repeating an experiment and getting the same results is not by itself enough, for whatever errors or self-delusions

may have occurred in the first instance might also be part of subsequent repetitions of the experiment (Hyman, 1977). That was precisely the case when, at the beginning of the twentieth century, the French physicist, Professor Blondlot, 'discovered' N-rays, an apparently new form of energy. He replicated his experiments many times, and indeed, a score or more of other scientists reported that they had confirmed the existence of N-rays in their own laboratories. Yet sceptical scientists were unable to replicate these results, and ultimately Blondlot's findings were shown to be a product of self-delusion (Alcock, 1981). The concept of replicability, to be useful, implies that researchers in general, provided that they have the expertise and equipment, should be able to reproduce the reported results, and not just those who are believers and enthusiasts.

Because parapsychologists have *never* been able to produce a successful experiment that neutral scientists, with the appropriate skill, knowledge and equipment, can replicate, some parapsychologists have gone so far as to argue that the criterion of replicability should not be applied to psi research because the phenomena are so different from the usual subject matter of science (Pratt, 1974). Yet, what a risky adventure it would be to yield to special pleading and relax the very rules of scientific methodology that help to weed out error, self-delusion and fraud in order to admit claims that violate the basic tenets of science as we know it!

Several of the papers in this Special Issue address the problem of replicability in psi research:

(1) My good and respected friend Adrian Parker acknowledges the highly problematic inconsistencies in parapsychology that reflect both failures to replicate and situations where some experimenters, but not others, can replicate a set of findings. Yet he does not take this to suggest that the Psi hypothesis might be wrong and the Null hypothesis correct, but instead views these irregularities as reflecting possible *properties* of the ostensible phenomenon, such as the *psi-experimenter effect* (discussed below). This is begging the question. When there has been a failure to replicate, it is not appropriate to engage in the circularity of assigning to this failure a label (psi-experimenter effect), and then implicitly suggesting the label as its explanation. Since there is no other way of defining or identifying the psi-experimenter effect, it has no *explanatory* value. Using it as a possible explanation only leads to a tautology: by substituting the definition of the psi-experimenter effect, one gets 'The failure to replicate may be a manifestation of "one researcher failing to replicate a finding that another researcher had made".' This circular reasoning excludes from the debate a possibly fruitful aspect of research, in terms of coming to understand the reasons, other than psi, that might account for the fact that different experimenters have obtained different results.

(2) With regard to ESP in the ganzfeld, Palmer concludes that, while he finds statistically significant departures from the Null hypothesis across the aggregate data bases that he has examined, 'the marked heterogeneity of results across experiments leaves doubt about the future replicability of the phenomenon outside parapsychology'.

(3) In their article, Sherwood and Roe examine attempts to replicate the well-known Maimonides dream studies that began in the 1960s. They provide a good review of these studies of dream telepathy and clairvoyance, but if one thing emerges for me from their review, it is the extreme messiness of the data adduced. Lack of replication is rampant. While one would normally expect that continuing scientific scrutiny of a phenomenon should lead to stronger effect sizes as one learns more about the subject matter and refines the methodology, this is apparently not the case with this research. They conclude: 'Overall, the Maimonides studies were more successful than the post-Maimonides studies but this may be due to procedural differences.' Indeed, this leads the authors to indicate that 'more recent work has concentrated on the question of whether consensus methods are superior to individual performance. With consensus judgement procedures, the responses from a number of individuals are combined to give a single judgement.' To the sceptic, this is a strange turn of events. The phenomenon of interest is the alleged ability of some individuals to paranormally receive information while they are asleep. Because research cannot demonstrate this clearly, the researchers choose to complicate the situation immensely by combining information from a number of subjects.

(4) Jeffers' article also bears directly on the question of replicability. Jeffers stands in lonely company as one of the very few neutral scientists who have empirically investigated the existence of psi phenomena. My first interaction with Jeffers is memorable to me. Jeffers, a physics professor at my university, was inspired by the work of Robert Jahn (e.g. Jahn, 1982), that purported to demonstrate the influence of the human mind on the output of a random event generator, and he decided to carry out his own psi experiments. His methodology was different from Jahn's (or indeed from other psi experiments) in that it investigated the possible effect of psi on the interference of light. He reasoned, and Jahn had agreed, that if Jahn's results were due to subjects' mental influence on quantum processes, then that same influence might be expected to affect the interference patterns produced when two beams of light are sent through narrow slits. In Jahn's work, a series of numbers appeared on a computer screen, the ultimate result of a quantum process, and subjects strove to affect the magnitude of those numbers. In Jeffers' work, a bar appeared on a computer screen, its length determined by a quantum process (fringe contrast in the interference pattern) and subjects attempted to influence the height of the bar. Thus, Jahn and Jeffers were both attempting to measure subjects' ability to influence quantum processes by mentation alone and, given that different methodology was used, were Jeffers' research to have produced significant results this would have added even more weight to Jahn's conclusions than would a straight replication. This is because Jeffers studied the same construct, or concept, from a slightly different angle, thereby making his research capable of producing convergent evidence, whereas a straight replication using exactly the same methodology might also reproduce any undetected errors and biases in the original.

Back to our initial meeting: Jeffers came to me at least a tad defiantly, requesting that I review his experimental design and offer any suggestions and

criticisms before he began his research. He stressed that I should not after the fact, were he to obtain data supporting the parapsychological interpretation, then argue that the experiment was not to be taken seriously because it had fallen methodologically short in some fashion. Thus began our relationship, which was to grow into the very positive one that it is today. I reviewed his experimental design, and I raised some reservations — the same reservations that I had written about (Alcock, 1990) with regard to Jahn's work. While so far as I am aware, Jahn's group never paid any heed to my comments, Jeffers incorporated changes that satisfied all my concerns. As Jeffers reports in his paper, his research findings give no support to the Psi hypothesis.

Jeffers' research makes a very important contribution to the study of putative psi phenomena, in my opinion, for the following reasons:

1. It was carried out by a neutral scientist who approached the subject with great interest and motivated by the possibility that Jahn may really have discovered something very important — the influence of human mentation on random physical processes. This should be an ideal condition for producing the desired results: Jeffers was very much open to the possibility of psi and was motivated to find it.
2. The research began with the full approbation of both proponent and sceptic. Jeffers' had the full-fledged support of Jahn himself and, as noted above, I fully supported the appropriateness of the revised methodology that he employed. Had he produced positive results, Jahn no doubt would have viewed this as a significant conceptual replication of his own work by a neutral scientist, and I in turn would have had to admit that the research was done carefully and correctly, and that I had no basis for rejecting it on methodological grounds.

However, when Jeffers' research did not produce results supportive of the Psi hypothesis, other researchers in the area dismissed it, and now it receives virtually no attention from parapsychology at all. (To be precise, his article discusses two kinds of experiments, one single-slit and one double-slit. The results of the single-slit experiment, carried out at York University, were null. There were two sets of double-slit experiments, one conducted at York University and one carried out in Jahn's laboratory at Princeton. The York experiment produced a null outcome, while that at Princeton produced 'marginal' significance ($p = 0.05$), which Jeffers views, as do I, as unconvincing). This neglect of Jeffers' research is most unfortunate. Although his data, as reviewed in his current paper, is in line with the Null hypothesis, the fact that it is now ignored within parapsychology is another instance of not giving the Null hypothesis a fair chance.

Incidentally, Jahn's laboratory more recently collaborated with researchers at two German universities to attempt a carefully controlled replication of the basic claims of Jahn's research group. The result? Neither the researchers at Jahn's lab nor those in the two German universities found anything of significance with regard to the hypotheses under test (Jahn *et al.*, 2000). They did, however, on a *post-hoc* basis — as is so often the case in parapsychology —

find some 'anomalies' in the patterning of the data which they argue call for more sophisticated experiments and theoretical models in order to understand 'the basic phenomena involved'. Again, failure to confirm predictions does not, in their view, give strength to the Null hypothesis. By post-hoc data snooping, a success of sorts can always be wrestled away from the jaws of the Null.

In sum, parapsychologists have never been able to produce a demonstration that can be reliably replicated by researchers in general, and failures to replicate are either ignored, explained away or interpreted as evidence for the existence of arbitrary properties of psi, as is discussed below.

4. Multiplication of entities

Despite William of Ockham's exhortation that one should not increase the number of entities required to explain a phenomenon beyond what is necessary ('Ockham's Razor'), parapsychology has unabashedly invented a number of such entities by way of explaining away failures to produce consistent and replicable data. For example:

1. As touched on earlier, if only some researchers can obtain an effect — and then only some of the time — while other researchers using identical methods cannot, this is taken, not as lending support to the Null hypothesis, but as a manifestation of a property of psi — the *psi-experimenter effect.* This 'effect' supposedly occurs because some experimenters, perhaps because of their own psi abilities, are conducive to the production of psi in experiments, while others are not.

   Smith's article in this Special Issue provides a good overview of the enduring problem of the experimenter effect in parapsychology, but his analysis also indirectly serves to demonstrate the problem that I am addressing. While acknowledging the issue of replication in parapsychology, Smith argues that 'replication difficulties in parapsychology may be due, at least in part, to psi-related experimenter influences'. He recognizes that this view is difficult from the point of view of science because it suggests that 'it is only those researchers who believe that psi exists that are likely to be able to replicate positive results'. Nonetheless, as he reflects upon this problem, Smith's optimism is not diminished and he argues: 'the scientific approach adopted by psi research has so far achieved some limited success in identifying factors associated with obtaining positive results in psi experiments, and it is my view that it is such an approach that is likely to reveal more of these factors in future research. Only when we have a much more detailed recipe for success can more consistent levels of replication be expected.' Thus, while aware of the problem he sidesteps it.

   Parker also addresses this subject, and states that 'experimenter effects and psi-conduciveness are every bit as integral part of the phenomena being studied as, say, placebo effects are in psychological treatment'. The problem is that the 'experimenter effect' is really only a lack of consistency, a lack of

general replicability, which itself is more in line with the Null hypothesis than anything else. There is no reason, no justification, to engage in further multiplication of explanatory entities, to use Ockham's language. What we have here is a failure to replicate. Period. The psi-experimenter effect provides the ultimate Catch-22: if you find the psi effect you are looking for, well and good. If you do not find it, this might be because of the experimenter effect, and so this too could be a manifestation of psi!

2. The *sheep–goat effect* refers to the observation that believers in psi are more likely than non-believers to demonstrate evidence of psi in an experiment.
3. If subjects fail to obtain the above-chance scores predicted in a psi experiment, that is not taken as lending weight to the Null hypothesis. Instead — so long as they fail miserably enough that their data deviate statistically significantly in the non-predicted direction, then this is taken as support for the Psi hypothesis, and another 'effect' — the *psi-missing effect* is invoked, allowing the interpretation that the miserable failure was indeed a success.
4. If a 'gifted' subject scores well in early trials but then, as is so often the case, scores only at a chance level later, this is not taken as support for the Null hypothesis. Instead, it is taken as evidence for another 'property' of psi — the *decline effect*. Thus, failure is often interpreted as a kind of success, as an indication of the weird properties that this elusive psi possesses.

I note that one such 'effect', at one time well-known within parapsychology, appears to have quietly disappeared. I am referring to the *quartile-decline effect*, much discussed by the pre-eminent parapsychologist Joseph Banks Rhine, and so-named because it was noted that when subjects' scores were recorded in two columns to a page, there was often a significant decline in subjects' success if one compared the scores in the upper left-hand quadrant of the page to those in the lower right-hand quadrant. While such an 'effect' always struck sceptical observers as somewhat convenient and arbitrary, it was touted as again suggesting some strange property of psi.

Indeed, the very fact that it has proven so difficult to produce a reliable demonstration of a psi phenomenon has led some researchers to think of this general elusiveness not as something in line with the Null hypothesis, but rather as another property of psi. Parker's paper speaks to this: 'For whatever reason the phenomena appear to have an elusiveness as a defining characteristic that makes them intrinsically difficult to capture in the laboratory in a stable, predictable and controllable fashion.'

Note that none of these so-called effects are anything other than arbitrary, post-hoc labels attached to unexpected negative outcomes. The employment of arbitrary *post hoc* constructs to explain away failures and inconsistencies in the data is a serious problem when one considers the scientific status of parapsychology. The Null hypothesis is not given a fair chance when data that are consistent with it are explained away in this manner.

5. Unfalsifiability

Obviously, the use of such 'effects' as those just discussed serves to make claims about psi essentially unfalsifiable, for any failure to produce the predicted effect, or any inconsistency in the data, can be explained away in terms of one or another of them. Failure to produce data consistent with psi has never been taken as providing weight to the null hypothesis.

Falsifiability is an important concept in science, especially when highly unusual claims are made. Science did not ignore Roentgen's rays just because they did not fit in with what was known at the time. On the other hand, science did not ignore Blondlot's rays (N-rays) either. The former turned out to be a highly replicable phenomenon that demanded changes in physical theory to account for it. The latter, despite numerous independent 'replications' initially, turned out to be a figment of the imagination. This is why falsifiability is so important.

6. Unpredictability

This problem is also related to the replication difficulty. Parapsychologists cannot in general make predictions before running experiments and then confirm them. Yet, as discussed earlier, even if predictions are not confirmed, researchers often point to some apparent irregularity in the data that suggests, *post-hoc*, that some other psi event occurred.

Yet, if psi is real, one might expect that psi manifestations would be predictable, as least to some extent. With the vast amounts of data that parapsychologists typically collect, it would be straightforward enough to calculate the number of datapoints needed to obtain an effect size of an arbitrary magnitude, and then rerun the study with that number of data points, and find the predicted effect if it is there. It never works out that way. This has led Palmer to admit to 'what appears to be an intractable problem in parapsychology. Until we can predict such outcomes ahead of time, the establishment of lawful relationships still evades us.' This unpredictability, I must point out, is what one would expect to find if the Null hypothesis, rather than the Psi hypothesis, obtains. If the Null hypothesis is true, if there is no such thing as psi, then 'significant results' occur from time to time because of a concatenation of chance factors, flaws in the experimental design, and so on. In such a case, one would not expect any lawfulness in the data, and one would not be able to predict what should occur in the next experiment based on what has happened in the last.

7. Lack of progress

Not only is there a problem of general inconsistency in the data, as discussed above, there has not been any real improvement in this situation over time. Despite the use of modern random event generators and sophisticated statistical analyses, parapsychologists are no closer to making a convincing scientific case for psi than was Joseph Banks Rhine back in the 1930s. There has been no growth in understanding. Psychic phenomena, if they exist, remain as mysterious as

ever. No consistent patterns have emerged. Effect sizes do not grow over time as a result of refinements in methodology. No well-articulated theory supported by data has been developed. Indeed, rather than producing a gradual accumulation of knowledge and an evolution of better and better methodology, every decade seems to spawn some new methodology or paradigm or research programme that offers promise of the long-awaited breakthrough, but that gradually loses its glitter. The famous Rhine experiments (e.g. see Rhine *et al.*, 1966/1940) are no longer held up as strong evidence for the Psi-hypothesis. Soal's research (e.g. Soal and Bateman, 1954), once trumpeted, is now forgotten, and for good reason. Targ and Puthoff's remote-viewing experiments (e.g. Targ and Puthoff, 1974) , which showed early promise, now are virtually ignored, again for good reason. The Maimonides research has been difficult to replicate, as Sherman and Roe point out. Jahn's research group at Princeton continues its efforts (e.g. Jahn *et al.*, 2000), but its impact is minimal within modern parapsychology, partly due to methodological problems identified by other parapsychologists and critics alike. There has been no real growth in understanding or in the ability to isolate the putative phenomena over time. New research strategies seem to 'fret and strut their hour upon the stage' and then are heard little more.

8. Methodological weaknesses

Given that psi is defined negatively, and can only assumed to have been present if all possible normal explanations can be ruled out, critics of parapsychology are naturally inclined to look for flaws in the experimental design and execution of research that would account for whatever positive effects parapsychologists have adduced. Of course, this quest is hampered by the fact that experimental reports will only rarely capture sources of error of which the experimenter was oblivious, and so it is not always possible in the first instance to find normal sources of putative psi effects based on the write-ups alone. The nub of the debate between sceptic and proponent is most typically the adequacy of the methodology. I think it fair to say, and I suspect that both Parker and Palmer would agree with me on this, for they have been strong methodological critics of much parapsychological research themselves, methodological weaknesses have, in a large number of studies, vitiated the claim to have demonstrated something paranormal. However, some parapsychologists have argued that even when errors and weaknesses are found, the onus is on the critic to show that the error could have produced the observed effects. That argument is not persuasive however, for the onus is always on the researcher to demonstrate that he or she has done the experiment well, and flaws in design or procedure show that it was not done well, and that perhaps other less obvious methodological problems have also been a factor. The answer is simply to run the experiment again, doing it right this time. That is what is expected in mainstream science. The problem for parapsychology, however, is that the difficulty in replication means that it may not be possible to get the same results a second time, whether the methodology is cleaned up or not.

However, are sceptics too intent on finding methodological flaws and in so doing failing to see the phenomenon of interest? One must, of course, be careful not to throw the baby out with the bathwater when one approaches data that do not fit in with the contemporary scientific worldview. There are many examples were the baby was thrown out, only to be rediscovered years later, crying out for attention. Mesmer argued that he was 'curing' hysterics by means of animal magnetism. Mainstream scientists of the day who were charged with the evaluation of his claims demonstrated that his explanation was wrong, that when the metal rods that he used in his procedure were secretly removed, this made no difference to the outcome, and his patients still responded positively to the procedure. Mesmer would not back down; he stood by his theory of animal magnetism, and the necessity of the metal rods. The scientists would not back down; they stood by their findings that magnetism had nothing to do with it. As a result, Mesmer's clinic was shut down, and both sides in the dispute missed a wonderful opportunity to discover and explore what we now call 'hypnosis'.

I agree with parapsychologists when they declare that if a single instance is known in which 'action at a distance' occurred, we at least know that contemporary science does not encompass the whole story about nature. Yet, if the observed action at a distance is not replicable, then it is questionable whether it has really been demonstrated to occur. Indeed, it is important to remember that hypnosis was ultimately 'discovered', though its true nature remains subject to some debate even today. While not everyone appears to be susceptible to hypnosis, just about anyone can quickly learn to produce hypnosis in susceptible subjects simply by following a standard script. In comparison, over a century of parapsychological inquiry has as yet failed to produce a publicly replicable demonstration of psi, and that despite its long history, parapsychology still lacks the evidence it needs to be placed before the scientific community for judgment. (Hyman, 1977; 1985; 1989). At some point, it seems justifiable to presume that there may not be a baby in the bathwater, and that the Null hypothesis is correct!

9. Reliance on statistical decision-making

Because of the failure to be able to produce a straightforward demonstration of psi ability, such as might be the case if a psychic could reliably predict winning lottery numbers, or if, as Gardner (1957) suggested many years ago, a psychic could cause a fine needle, which is carefully balanced on another needle and housed under a Bell jar from which the air had been evacuated, to rotate, parapsychologists at the more scientific end of the spectrum came to depend more and more upon statistical analyses to demonstrate their putative phenomena. With such an approach, subjects make guesses or make mental attempts to influence random event generators, and then their success or failure is judged by a statistical comparison with what would be expected by chance alone.

Statistical analysis was applied first in psychological research as a means of protecting the researcher against error. It allowed the researcher to evaluate the likelihood that his or her results, no matter how strong the data appeared to the

naked eye, could have occurred by chance alone. In recent years, such analysis has been employed to do much more than simply provide guidance about the likelihood that particular data may well have arrived by chance alone. Powerful statistical techniques now exist for finding patterns in data that elude the naked eye, and this provides an important tool for researchers in many domains. Thus, statistical analysis originally helped cool our ardour about what appeared to be meaningful effects in the data, whereas now, those statistical tools are used to find significant effects that we would not otherwise detect. Now, in modern parapsychology (and, alas, in mainstream psychology as well to some extent), statistical analyses are being used to define and defend the importance of differences so small that they would have carried no interest to researchers of a century ago. If subjects score at a rate of 51% when the chance rate is 50%, it is unlikely that anyone would have taken any notice a century ago. Now, provided the sample size is large, such a small difference may well be 'statistically significant'.

There is no reason in principle that such analysis should not also be used in parapsychology, but there is an important difference in the way that it is used in that field. In regular science, statistics are used either to look for covariation amongst well-defined variables, or to evaluate whether a given measurement is affected by the presence or absence of an 'independent' variable. However, in parapsychology, there are no well-defined variables, and there is no way of controlling whether psi (if it exists) is present or absent, and so the statistical process is used, not to evaluate the effect of one or more variables on other measurable variables, but as a basis for inferring the presence of psi itself. One begins with the assumption that a particular mathematical distribution describes the probability distribution of outcomes of a randomly generated event. A subject in some way tries mentally to influence the distribution of outcomes (even if he or she knows nothing about the nature of that distribution, or about the generator that produces it, or even where the generator is physically located). If the outcomes depart from the theoretical distribution to a significant extent, this is taken as evidence that a psi influence caused the departure.

Any such statistically significant departure is viewed as an 'anomaly' relating to psi, and thus is viewed as support for the Psi hypothesis. However, statistical significance tells us nothing about causality. If a person tries to guess or 'intuit' what number will come next in a randomly generated sequence, and succeeds better than one would expect by chance, that tells us absolutely nothing at all with regard to why such results were obtained. The departure from chance expectation could be due to any number of influences — a non-random 'random generator', various methodological flaws, or . . . Zeus. (I could posit that Zeus exists and likes to torment parapsychologists, and thereby gives them significant outcomes from time to time, but does not allow replication outside parapsychology. The significant outcome would provide as much support for my hypothesis that Zeus exists as it does for the Psi hypothesis that the human subject's volition caused the results.)

Joseph Banks Rhine, whose psi research was motivated in part by the desire to find scientific evidence for post-mortem survival, passionately believed in the

scientific method, and consequently he shepherded parapsychology into the laboratory and into the research paradigms favoured by experimental psychologists — studies with specific targets, controlled conditions and statistical analysis of data. It was at that juncture that, despite the admirable effort to harness emerging social science technology, formal parapsychology began to lose touch with the very experiences that originally motivated its pursuit. Rather than focusing on conditions that seem to be conducive to paranormal experiences such as telepathy in everyday life, and then seeking at first to understand the experience in terms of normal psychology, these laboratory studies focused only on one explanation of such experiences — the notion that somehow there is a transfer of information that does not involve any known sensory apparatus or energy. Thus, the laboratory approach involved trying to 'send' information from one brain to another, or trying to 'read' objects hidden from view, or trying to predict the outcome of the roll of dice before they are thrown. Note that such activities have virtually nothing to do with human experiences that seem to many to be paranormal. Worse, verification of the supposed success of the psi task became a statistical one. No longer was it a question of whether a person had dreamed of his father's funeral in detail, not knowing that miles away his father had died, but rather, what is the series of cards that will next be turned up?

10. Problem of theory

That quintessential investigator Sherlock Holmes once opined: 'It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.' This is also good advice when it comes to theorizing in parapsychology. The database does not at this time justify the development of explanatory theory, for, as I have discussed above, it is far from clear that there is anything to explain. Notwithstanding the absence of good evidence, there have been many attempts to develop theories to explain putative psi phenomena, among them the Conformance Behaviour model (Stanford, 1990); Decision Augmentation Theory (May, Utts and Spottiswoode, 1995); a teleological (goal-seeking) theory (Schmidt, 1975); a quantum mechanical theory (Walker, 1984); the Thermal Fluctuation Model (Mattuck, 1982) that proposes that the 'mind' somehow alters the outcome of an event by manipulating the thermal energy of molecules; and, as described in this Issue, Pallikari's statistical balancing theory (discussed below). Such theorizing in the absence of reliable data, especially when it attempts to interpret quantum mechanical theory in such a way as to accommodate psi, lends an unjustified patina of scientific respectability to parapsychology, especially in the eyes of those who are outside the world of physics.

Jeffers' paper critically discusses the argument often heard within parapsychology that quantum physics in some way or another can accommodate/explain psi. Pallikari, on the other hand, begins with the assumption that psychokinesis occurs, and her analysis concludes that it only operates at a micro-level, and therefore does not show up at the macro level. She proposes a theoretical approach

to understanding such micro-PK, and inherent in this notion is the idea of statistical balancing in the long run, so that macro-PK will not be observed. While I admire Pallikari's efforts, they are premature, for the problem remains that to date there are no substantive empirical data to justify such theorizing. Of course, her conclusion that psychokinesis does not show up except at the micro level is at variance with what many other parapsychologists have claimed to have observed at the macro level.

11. Failure to jibe with other areas of science

A major criticism of parapsychology is that it fails to jibe with other areas of science. The late neuropsychologist Donald Hebb (1978) once commented that if parapsychology is right, then physics and biology and neuroscience are horribly wrong in some fundamental respects. He went on to say that science has been wrong before, but that parapsychology would need very strong evidence if it was going to be able to challenge successfully the current state of knowledge in mainstream science. For example, psi influences, unlike any known energy, are invariant over distance. Time produces no barrier either, apparently, for such influences are said to be able to operate backwards and forwards in time. If the 'out-of-body experience' is a psi effect, then it would apparently demonstrate that the complex mechanisms of the brain, while extremely vulnerable to disruption or total destruction as a result of disease or injury, are apparently unnecessary for perception or cognition in the out-of-body individual. To be fair, some parapsychologists have argued that their data tends to support the idea that the brain does indeed process incoming psi. Yet such processing is not a simple matter, for as Beyerstein (1987) noted, in pointing to the profound implications that psi would have for the neurosciences. He pointed out that perception, memory and emotion involve extremely complex neurochemical configurations that are the result of the spatiotemporal integration of activity in millions of widely-distributed neurons and their internal components. *Extrasensory* perception would by definition bypass the activity of peripheral receptors and nerves that normally determine these central electrochemical configurations. To experience the emotion or the percept, then, any hypothetical 'psi signal' would have to produce the corresponding central electrochemical configurations directly, which would involve influencing the internal chemical processes of millions of neurons in the correct sequences and in the appropriate anatomical pathways. This, in the view of neuroscientists in general, is highly unlikely. Yet while there are attempts to interpret physical theory in such as way as to accommodate psi (e.g. Pallikari in this Issue), parapsychologists appear disinterested in the contradictions between parapsychology and neuroscience (Kirkland, 2000).

On the other hand, failure to jibe with other areas of science is in a very real sense the *sine qua non* of parapsychology. As discussed earlier, something is only considered paranormal if it defies current scientific models of reality.

12. Disinterest in competing hypotheses

Unfortunately, the focus in parapsychology seems to be more on finding the anomaly than on explaining the experience. The retreat into the laboratory has led to a focus on statistical deviations that thoroughly distracts researchers from seeking other, more prosaic, 'normal' explanations, for psychic experiences. That is too bad, for as I have indicated earlier, such experiences warrant study in their own right, regardless of whether there is any need to appeal to paranormal explanations. I have many times argued (e.g. Alcock, 1981) that even if psi does not exist, we should still expect that most people will have experiences in their lifetimes that *seem* to yield to no other explanation than a psychic one, and I have explained just why that should be so and have offered explanations as to how various paranormal experiences can be the result of normal (and sometimes abnormal) brain function. Many other psychologists (e.g. Beyerstein, 1987–8; 1988; Blackmore, 1982; Marks, 2000; Neher, 1990) have also provided substantial and detailed explanations with regard to how normal and abnormal psychological processes are capable of producing all the elements of paranormal experiences. This information should be of great interest, one might think, to parapsychologists, but the question of normal causality seems usually to be dismissed out-of- hand as being unimportant to the study of the paranormal.

Several of the articles in this Special Issue address some of the factors that might explain why some people believe that they have witnessed or experienced paranormal phenomena even if they have not.

1. Ultimately, 'real-life' accounts of paranormal experiences, the very sort of accounts that led to parapsychological research in the first place, rely on processes of perception, interpretation and memory. French provides a good discussion of how such factors as hallucinations, imagery, suggestibility, dissociative tendencies and unreliable memory may be the well-spring of many such accounts.
2. Dean and Kelly are the preeminent critics of astrology and its claims, and they note that astrology actually seems to 'work' for many people — perhaps many millions of people — around the globe. Thus, there is widespread belief in astrology just as there is in parapsychology, and there is a similarity between the two, in that both essentially involve correlations between two events, and the imputation of causality. In parapsychology, one 'wishes' or 'guesses' — either in real life or in the laboratory, and then any significant correlation between wish/guess and outcome is taken as evidence of causality. For example, in a psychokinesis study, a subject wishes to produce high numbers in the output of a random number generator, while in a telepathy study, the subject essentially guesses at what the sender is sending. Similarly in astrology, the astrologer produces a description of one's future, and to the extent that it seems to correspond with what happens later, it is taken to support the notion that the position of the stars at birth is related causally to later events in one's life. Dean and Kelly show that astrology seems 'to work' because of the cognitive errors that individuals make in reacting to their

horoscopes — the fallacy of personal validation. A similar explanation can be applied to the readings offered by psychics. What is particularly important to the discussion of parapsychology in their article is the pervasive extent to which people can come to strongly believe in a demonstrably false system of causality.

3. Brugger and Taylor adduce evidence that supports their contention that believers in paranormal phenomena more readily perceive meaningful associations in random stimuli than do disbelievers, and argue that believers develop an 'illusion of control', perceiving a causal relationship between their actions and environmental events that produces a strong belief in a paranormal causation of the event. They further argue that believers tend not to test alternative hypotheses. To me, their paper is a particularly important one, for it shows the way towards understanding how the exigencies of both everyday life and the parapsychology laboratory can be expected to generate strong impressions that something 'psychic' has occurred.

These are some of the reasons that I would urge caution in one's approach to parapsychological claims. However, no doubt parapsychologists would argue that I am being unfair and overly negative.

This leads me to another question: *Has mainstream science been unfair?* Parker contends that mainstream science has not given parapsychology a fair hearing. I respectfully disagree. I have detailed elsewhere (Alcock, 1987; 1990) how conventional science and mainstream psychology have actually provided numerous opportunities over the years for parapsychologists to bring their work to a larger scientific audience. Indeed, when the American Society for Psychical Research was founded in 1885, its membership included several prominent psychologists of the day, most of whom eventually left the organization when they failed to find any evidence of psychic phenomena. Again, in the early part of the twentieth century, other prominent scientists and psychologists were open to the study of parapsychology, and some undertook studies of their own but gave up when their efforts failed to produce results. In the 1930s, not only did the American Psychological Association sponsor a round-table discussion of parapsychology, but a 1938 poll found that 89% of psychologists at that time felt that the study of ESP was a legitimate scientific enterprise (Moore, 1977). Various scientific publications over the years, including prestigious psychological journals such as *Psychological Bulletin*, have brought parapsychological research and views to the non-parapsychological scientific community. Indeed, between 1950 and 1982, more than fifteen-hundred parapsychological papers were abstracted in the American Psychological Association's *Psychological Abstracts* (McConnell, 1977). Nonetheless, mainstream science continues to reject parapsychology's claims. In my mind, this is not because of some unfair bias, but simply because parapsychologists have not been able to produce data that persuade the larger scientific community that they have a genuine subject matter to study.

This lack of acceptance by science no doubt creates cognitive dissonance on the part of those parapsychologists who are convinced that they do have real

phenomena. This dissonance can be resolved either by assuming that the exclusion from the halls of mainstream science is unfair and unjustified, or that there is some reason other than lack of persuasive data that underlies the rejection. As an instance of the latter, the prominent parapsychologist Charles Tart once wrote that sceptical scientists may be unconsciously so afraid of their own psychic abilities that they have to attack any evidence that might provoke knowledge of their own ability (Tart, 1982; 1984). Parker, in this Issue, argues that perhaps sceptical psychologists do not really want to resolve the issue about the reality of psi for fear of the 'unwanted implications' for psychology if it were shown that psi really does exist. He may be correct, but I doubt it. In my many years in the field of psychology, I have never detected anything other than simple disinterest in parapsychology from the vast majority of psychologists. They simply assume that psi phenomena have never been shown to exist. On the other hand, I am certain that were there suddenly to be produced compelling evidence for the reality of psi, parapsychologists would be knocked over in the stampede by experimental psychologists to explore an exciting new area of research.

Can the psi question be resolved? Parker argues that the technology now exists that would allow a resolution of the question of whether psi exists, and that it would be relatively straightforward to resolve the question, were it not for a lack of funding from mainstream science. He also states that parapsychology might turn out to present genuine phenomena — or, it could turn out to be based on a mixture of fraud, artefact and subjective validation.

I would certainly applaud any effort and investment directed at resolving the psi issue, but I do not think that it is really possible to resolve it, unless of course compelling and replicable demonstrations of the existence of psi are forthcoming. I do not believe that parapsychologists give the Null hypothesis a proper chance, and I cannot conceive of any research that could serve to persuade parapsychologists that psi does not exist. It would be far easier, were good and reliable data available, to persuade sceptics of the reality of psi than to dissuade parapsychologists. What evidence can one produce with regard to 'disproving' the psi hypothesis? Certainly not carefully executed studies that fail to replicate, that fail to produce any evidence of a psi anomaly. Those are too easily explained away in terms of the 'experimenter effect' or simply ignored, as is the case with Jeffers' research. Finding prosaic explanations for a given data set may persuade parapsychologists that, in that particular instance, there was no evidence for psi, but what about all the other data sets yet to come? Parapsychologists can neither tell us under what circumstances psi, if it is real, does not occur, nor can they tell us how it would be possible to disprove its existence.

While some parapsychologists, as noted earlier, ascribe hidden motivations to the continued resistance of mainstream scientists to bring parapsychology into the scientific fold, I judge it unlikely that parapsychologists would under any circumstances abandon their belief in and pursuit of the paranormal. In fact, while Brugger and Taylor propose the joint collaboration of traditional parapsychology and neuroscience in the hope that findings from prospective research conducted by representatives of two apparently conflicting views will most likely be taken

seriously by both sides, they also foresee what many parapsychologists would consider to be an unacceptable downside: 'We thus anticipate that, although psi would vanish from the scene as a process of information transfer, it would live on as a phenomenon of subjective probability worthy of scientific investigation.'

Finally, even if one were to produce a set of circumstances that would lead some parapsychologists to abandon the psi hypothesis, parapsychology as a whole would carry on much as it always has, and the conclusions of those who left the field would be downplayed or ignored, just as were Blackmore's conclusions when she pronounced that she had become sceptical with regard to psi and was leaving the field, or Wiseman's as he had become more and more identified with the sceptical position (Wiseman, 1997). Of course, for those who appropriate for themselves the label 'parapsychologist', but do not really subscribe to the appropriateness of a scientific examination of psi in any case, any agreement by science-oriented parapsychologists that resolves the psi question in a negative direction would carry no weight at all.

Thus, the search for psi will go on for a long time to come, for I can think of nothing that would ever persuade those who pursue it that the Null hypothesis is probably true. Yet, as this search goes on, those of us who are sceptics should applaud and support the approach taken by parapsychologists who have contributed to this Special Issue — not because we agree with their conclusions, for we shall continue to scrutinize and, when appropriate, find fault with their methodology and challenge their interpretations — but because they share our belief in the power of the scientific method to reveal truth in nature. I do marvel at their tenacity, however, for they labour in search of psi despite a lack of the evidentiary and other rewards that are earned by mainstream scientists in their research. Yet, that being said, and as I have stated before (Alcock, 1985; 1987), I continue to believe that parapsychology is, at bottom, motivated by belief in search of data, rather than data in search of explanation. It is the belief in a larger view of human personality and existence than is accorded to human beings by modern science that keeps parapsychology engaged in their search. Because of this belief, parapsychologists never really give the Null hypothesis a chance.

Acknowledgements

I wish to thank Jean Burns and Anthony Freeman for their very helpful comments with regard to the draft version of this manuscript.

References

Alcock, J.E. (1981), *Parapsychology: Science or Magic?* (London: Pergamon).

Alcock, J.E. (1985), 'Parapsychology as a "spiritual science" ', in *A Skeptic's Handbook of Parapsychology*, ed. P. Kurtz (Buffalo, NY: Prometheus Books), pp. 537–69.

Alcock, J.E. (1987), 'Parapsychology: science of the anomalous or search for the soul?', *Behavior and Brain Sciences*, **10** (4), pp. 553–65.

Alcock, J.E. (1990), *Science and Supernature: A Critical Appraisal of Parapsychology* (Buffalo, NY: Prometheus Books).

Beyerstein, B.L. (1987), 'Neuroscience and psi-ence', *Behavioral and Brain Sciences*, **10** (4), pp. 571–2.

Beyerstein, B.L. (1987–8), 'The brain and consciousness: implications for psi phenomena', *The Skeptical Inquirer*, **12** (2), pp. 163–73.
Beyerstein, B.L. (1988), 'The neuropathology of spiritual possession', *The Skeptical Inquirer*, **12** (3), pp. 248–62.
Blackmore, S.J. (1982), *Beyond the Body: An Investigation of Out-of-Body Experiences* (London: Heinemann).
Gardner, M. (1957), *Fads and Fallacies in the Name of Science* (New York: Dover).
Hebb, D.O. (1978), Personal Communication, cited in Alcock, J.E. (1981), *Parapsychology: Science or Magic?* (New York: Pergamon).
Hyman, R. (1977), 'The case against parapsychology', *The Humanist*, **37**, pp. 37–49.
Hyman, R. (1985), 'A critical overview of parapsychology', in *A Skeptic's Handbook of Parapsychology*, ed. P. Kurtz (Buffalo, NY: Prometheus Books), pp. 3–96.
Hyman, R. (1989), *The Elusive Quarry: A Scientific Appraisal of Psychical Research* (Buffalo, NY: Prometheus Books).
Jahn, R.G. (1982), 'The persistent paradox of psychic phenomena — an engineering perspective', *Proceedings of the IEEE*, **70**, pp. 136–70.
Jahn, R., Dunne, B., Bradish, G., Dobyns, Y., Lettieri, A., Nelson, R., Mischo, J., Boller, E., Böösch, H., Vaitl, D., Houtkooper, J. and Walter, B. (2000), 'Mind/Machine Interaction Consortium: PortREG Replication Experiments', *Journal of Scientific Exploration*, **14** (4), pp. 499–555.
Keen, M., Ellison, A. and Fontana, D. (1999), 'The Scole Report', *Proceedings of the SPR*, **58**, pp. 150–392.
Kirkland, K. (2000), 'Paraneuroscience?', *The Skeptical Inquirer*, **24** (3), pp. 40–3.
Marks, D. (2000), *The Psychology of the Psychic* (Buffalo, NY: Prometheus Books).
Mattuck, R.D. (1982), 'Some possible thermal quantum fluctuation models for psychokinetic influence on light', *Psychoenergetics*, **4**, pp. 211–25.
May, E.C, Utts, J.M. and Spottiswoode, S.J.P. (1995), 'Decision Augmentation Theory: towards a model of anomalous phenomena', *Journal of Parapsychology*, **59** (3), pp. 195–220.
McConnell, R.A. (1977), 'The resolution of conflicting beliefs about the ESP evidence', *Journal of Parapsychology*, **41**, pp. 198–214.
Moore, L. (1977), *In Search of White Crows* (Oxford: Oxford University Press).
Neher, A. (1990), *The Psychology of Transcendence* (New York: Dover).
Pratt, J.G. (1974), 'In search of a consistent scorer', in *New Directions in Parapsychology*, ed. J. Beloff (London: Elek Science).
Rhine, J. B., Pratt, J.G., Stuart, C.E., Smith, B.M. and Greenwood, J.A. (1966), *Extra-Sensory Perception after Sixty Years* (Boston: Bruce Humphries). (Original work published 1940.)
Schmidt, H. (1975), 'Towards a mathematical theory of psi', *Journal of the American Society for Psychical Research*, **69** (4), pp. 301–20.
Soal, S.G. and Bateman, F. (1954), *Modern Experiments in Telepathy* (New Haven, CT: Yale University Press).
Stanford, R.G. (1990), 'An experimentally testable model for spontaneous psi events', in *Advances in Parapsychological Research*, **6**, ed. S. Krippner (Jefferson, NC: McFarland and Co), pp. 54–167.
Targ, R. and Puthoff, H.E. (1974), 'Information transfer under conditions of sensory shielding', *Nature*, **252**, p. 602.
Tart, C. (1982), 'The controversy about psi: two psychological theories', *Journal of Parapsychology*, **46**, pp. 313–20.
Tart, C. (1984), 'Acknowledging and dealing with the fear of psi', *Journal of the American Society for Psychical Research*, **78**, pp. 133–43.
Walker, E.H. (1984), 'A review of criticisms of the quantum mechanical theory of psi phenomena', *Journal of Parapsychology*, **48**, pp. 277–332.
Wiseman, R. (1997), *Deception and Self-Deception: Investigating Psychics* (Buffalo, NY: Prometheus Books).

Part V
Experimenters' Personal Perspectives

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

[25]

The Elusive Open Mind: Ten Years of Negative Research in Parapsychology

What does a psychologist who's had an extraordinary experience do? Sets up a research program to test for psi. The lessons are surprising.

Susan Blackmore

EVERYONE THINKS they are open-minded. Scientists in particular like to think they have open minds, but we know from psychology that this is just one of those attributes that people like to apply to themselves. We shouldn't perhaps have to worry about it at all, except that parapsychology forces one to ask, "Do I believe in this, do I disbelieve in this, or do I have an open mind?"

The research I have done during the past ten or twelve years serves as well as any other research to show up some of parapsychology's peculiar problems and even, perhaps, some possible solutions.

I became hooked on the subject when I first went up to Oxford to read physiology and psychology. I began running the Oxford University Society for Psychical Research (OUSPR), finding witches, druids, psychics, clairvoyants, and even a few real live psychical researchers to come to talk to us. We had Ouija board sessions, went exploring in graveyards, and did some experiments on ESP and psychokinesis (PK).

Susan Blackmore is with the Brain and Perception Laboratory, University of Bristol, Bristol, England. *This article is based on her presentation at the 1986 CSICOP conference at the University of Colorado in Boulder. Her book* Adventures of a Parapsychologist *has recently been published by Prometheus Books.*

". . . I had an experience that was to have a lasting effect on me—an out-of-body experience."

Within a few weeks I had not only learned a lot about the occult and the paranormal, but I had an experience that was to have a lasting effect on me—an out-of-body experience (OBE). It happened while I was wide awake, sitting talking to friends. It lasted about three hours and included everything from a typical "astral projection," complete with silver cord and duplicate body, to free-floating flying, and finally to a mystical experience.

It was clear to me that the doctrine of astral projection, with its astral bodies floating about on astral planes, was intellectually unsatisfactory. But to dismiss the experience as "just imagination" would be impossible without being dishonest about how it had felt at the time. It had felt quite real. Everything looked clear and vivid, and I was able to think and speak quite clearly.

You can imagine the intellectual conflict I experienced (and of course I had no idea it was only a prelude to far worse mental conflicts!). The psychologists and physiologists who were teaching me made quite different assumptions about human nature from those made by the people I met through the OUSPR. The latter, for the most part, assume that there is "another dimension" to man, that we can communicate directly mind to mind, that there are "other worlds" waiting to be explored in altered states of consciousness, and even that consciousness is separable from its physical home and might survive the death of its body. The conflict was a challenge to me and I conceived the objective (I think naively, rather than purely arrogantly) of proving my teachers wrong, or at least showing that psychologists were closed-minded in ignoring the most important of human potentials—the paranormal.

Even at that very early stage I made a crucial mistake—or a series of crucial and related mistakes. First, I assumed that all these odd and inexplicable things—ESP, PK, OBEs, mystical experiences, ghosts, poltergeists, and near-death experiences—were related and that one explanation would do for all. Second, I assumed that there had to be a paranormal explanation—that we were looking for psi. Third (and I don't know whether this was just cowardice or an attempt at being sensible for a change), rather than launching straight into what really interested me—the OBE—I thought it was more "scientific" to begin with psi. After all, there had been research done on ESP and PK and, though generally rejected, it had some basis in scientific research. It seemed far easier, and safer, to start there. I didn't notice what I was doing. I can only point it out with the benefit of hindsight. I just took psi to be the key to the mysteries and wanted to study parapsychology.

The first thing I did was to develop my own theory of psi. This theory involved the notion that psi and memory are aspects of the same process,

that memory is a specific instance of the more general process of ESP. Eventually I got a place at Surrey University to do a Ph.D., and it was then that I set about testing my theory.

While I was at Surrey I was lucky enough to be given the chance to teach a parapsychology class. It attracted more than a hundred students, so I had plenty of subjects for my experiments. I began three kinds of tests. First, I predicted a positive correlation between ESP and memory. That is, if memory and ESP are aspects of the same process, then the same people should be good at both of them. I did many tests of this kind (Blackmore 1980a). Second, I predicted that the best target materials for ESP should not be those that are easy to perceive, but those that are easy to remember. I did a series of experiments with different target materials (Blackmore 1981a). Third, I predicted that the errors and confusions made in ESP should more closely resemble those made in memory than those made in perception. I had high hopes for this method since the study of errors has always been so useful in psychology, for example, in the study of visual illusions. I also did many experiments to test this (Blackmore 1981b). However, the only noteworthy thing about all of the results was the number that were not significant.

After a long series of experiments I had no replicable findings and only a large collection of negative results. Clearly they could not answer my original questions, nor test my special theory. Some of you may already be protesting: What an idiot. Why didn't she just give up and do something useful instead? But I would have responded: *This* could be useful! If ESP exists, it could be one of the most important findings for science; and in any case you can never tell in advance what research will be useful in the end. You may also be thinking, as many people said at the time: "Oh but this is just what you'd expect. She has only shown that there is no psi." But of course I hadn't done that, and couldn't do that. No amount of negative results can prove the nonexistence of psi. Psi might always be right around the next corner, and there were plenty of corners to look around.

"After a long series of experiments I had no replicable findings and only a large collection of negative results."

There were also plenty of parapsychologists eager to suggest corners I had not yet turned and reasons why my experiments had not worked. And I was eager to carry on the search. Some said it might be the subjects; students are notoriously not the best ones. So, instead of testing my class, I tested people who came to me with claims of special powers. I tried to design experiments that would test what they claimed to be able to do and that would allow me to impose sufficient controls. In some ways this upset me more than anything, because I met lots of genuine and well-meaning people who were convinced they could communicate by telepathy, or find underground pipes or hidden

water, until they tried to do it under conditions that ruled out normal sensory information. Then they, and I, were always disappointed.

"There were plenty of parapsychologists eager to suggest corners I had not yet turned and reasons my experiments had not worked."

Then I tried using young children as subjects. At that time, Ernesto Spinelli was getting outstandingly good results with preschool children in ESP tests (Spinelli 1983). So I set about designing experiments to use a method similar to his (though not a direct replication) to test my memory theory. It was much harder work than the previous experiments, but much more fun. The children were three- to five-year-olds in playgroups, and they thoroughly entered into the whole idea, being convinced they could transmit pictures to one another. But the results were quite clear. The proportion that were "nonsignificant" was as high as before. The overall results were nonsignificant and so were the correlations with age (Blackmore 1980b).

Why? Spinelli had many suggestions. It could have been that I used colored pictures, while his were black and white; or that the sweets I used as a reward (based on someone else's previously successful experiments) were too well liked by the children and were disruptive; or that I simply didn't have the right personality and rapport with the children. I could only say that I seemed to get on well with the children, but perhaps this was not well enough.

Another suggestion was that the problem was not the subjects themselves, but the state of mind they were in during the experiments. At that time, the ganzfeld experiments were the "latest thing," and the results from Carl Sargent (1980) at Cambridge, and Chuck Honorton (1977) at Princeton, seemed impressive. So I set about doing a ganzfeld study. My subjects each had half of a ping-pong ball covering each eye, lay on a reclining chair, and heard only white noise fed through headphones. I wrote down everything they said. Then they had to look at four pictures and choose which one they thought the agent had been looking at.

I had for some months led an imagery training group, in which we practiced relaxation, guided imagery, and many imagery tasks adapted from Buddhist training techniques. For my ganzfeld study I chose ten test subjects from this group and ten control subjects.

This study taught me a lot. Being in ganzfeld is in itself an interesting experience. Images come pouring in, and it is tempting to imagine that you are picking them up from somewhere outside of yourself. I also had one very impressive experience in which I was subject and my brother was agent. I "saw" people fishing, lakes, mountains, and swiss chalets, and when I saw the targets I picked the correct one right away. It was an amazingly close hit. It set me to wondering whether I had at last found the key! However, in the

course of the experiment I saw many equally amazing correspondences, but to the wrong pictures. My remarkable hit rapidly disappeared among the chance scores.

This should have taught me something important, something I should have known all along; that is, one should not rely on subjective estimations of probability (see Blackmore and Troscianko 1985). One should rely only on the statistics, and they were telling me that there was nothing there. Of course I tried it again with my brother, but the second time it did not work. Overall the results were close to chance expectation.

Why did this study also fail? I had used trained subjects in psi-conducive conditions and a method others had found successful. The ultimate suggestion of most parapsychologists was that it was an experimenter effect—more than that, it was a psi-mediated experimenter effect. That is, either I was using my own negative psi or I had some kind of personality defect, or defect in belief, that suppressed the psi of other people. I was a psi-inhibitory experimenter, so that whatever I did I would always get negative results. I began to get the feeling that I had some creeping sickness. I was a failure, a reject; there was something in me that suppressed the true spiritual nature of other people. I tried not to let it upset me, but I must admit that there is something terribly unflattering about being labeled "psi-inhibitory"!

"I was told I didn't get results because I didn't believe strongly enough in psi, because I didn't have an open mind."

Well, what could I do about it? It is not entirely an untestable idea. But Sargent had already tested the personalities of successful and unsuccessful experimeters and found the successful ones to be extroverted, confident, non-neurotic, and so on. In fact I fitted the description quite well—except for my results.

The other key to my failures seemed to be belief. I was told that I didn't get results because I didn't believe strongly enough in psi, because I didn't have an open mind! But what could I do about that? I couldn't just change my beliefs overnight or test ten subjects while believing and another ten while not! I argued that in the beginning I had believed in psi and still had got no results, but I couldn't prove this against the counter-argument that I had never *really* believed at all.

However, I did have an idea. There were still things in which I did believe. I could test the Tarot. In my preoccupation with everything occult, I had been reading Tarot cards for about eight or nine years. They really did seem to work. People told me that I could accurately describe them using the cards, and this was, naturally, gratifying. I even thought it might have a paranormal basis. So I set about testing the cards, doing readings for ten people, keeping the procedure as close as possible to a normal Tarot reading,

but isolating myself, as the reader, from the subjects. They then had to rank all ten readings to see whether they picked their own more often than chance would predict (Blackmore 1983).

It worked! The results were actually significant. You can imagine my excitement—perhaps I had at last found something. Perhaps there was no psi to be found in the standard laboratory experiments, but something paranormal could appear when the conditions were closer to real life. But then I talked to Carl Sargent. He pointed out that all my subjects knew one another, and if they knew one another their ratings and rankings could not be independent. So I had violated an assumption of the statistical test I was using.

This seemed so trivial. Their knowing one another could not help them pick the right reading, could it? No it couldn't; but this meant that the estimate of probability was inaccurate—and, after all, the results were only marginally significant. So I repeated the experiment twice more with subjects who did not know one another. I expect you can predict the results I obtained—entirely nonsignificant.

You may choose to interpret these results in different ways. Some parapsychologists have claimed that the first experiment found genuine psi and that the later ones didn't summon the same attitude, the same novelty, the same enthusiasm, that made psi possible—or even that psi itself doesn't like being replicated. But I think I had finally reached a stage where I no longer felt it was worth pursuing such arguments. I chose this point to say: "I think that, however many more experiments I do on psi, I am probably not going to find it."

Now we finally come to the question: "What do these negative results tell us?" Of course the one thing they do not tell us is that psi does not exist. However long I went on looking for psi and not finding it they could not tell us that. But I found myself simply not believing in psi anymore. I really had become a disbeliever. Like one of those doors with a heavy spring that keeps it closed, my mind seemed to have changed from closed belief to closed disbelief.

But either way I suffered. There was mental conflict whether I believed or disbelieved. I had many questions. One was this: How far could I generalize these negative results? The situation was the converse of the normal situation in science when one gets positive results and has to ask how far they can be generalized. Here I had to ask whether my negative results applied only to those experiments carried out by me, at those particular times, or whether they applied to the whole of parapsychology. There is no obvious answer to that question. If one had replicability one could answer the question as one does in other areas of science. But without replicability it is impossible.

The next question was: How could I weigh my own results against the results of other people, bearing in mind that mine tended to be negative ones while everyone else's seemed to be positive ones? I had to find some kind of balance here. At one extreme I could not just believe my own results and ignore everyone else's. That would make science impossible. Science cannot

"I could not just believe my own results and ignore everyone else's. . . . I could not believe everyone else's results and ignore my own."

operate unless people generally believe other people's results. Science is, and has to be, a collective enterprise.

At the other extreme I could not believe everyone else's results and ignore my own. That would be even more pointless. There would have been no point in all those years of experiments if I didn't take my own results seriously. Indeed, it is a fundamental principle in science that one has to take notice of the results one finds.

So there is no right answer to how to weigh them up. And these problems are only aspects of the basic dilemma of parapsychology, which is whether to believe or disbelieve in the existence of psi. Either way, I suggest, one meets conflict.

In the believer's position one is saying: "I believe there is something negatively defined, defined as communication without the use of the recognized senses, or action without the use of the muscles of the body. I have faith that future experiments will find this thing, even though so far they have failed to produce a replicable effect." If one takes this position, then one not only has to accept the open-ended nature of the search but also has to face up to the mounting negative results.

But what about the disbeliever's position? The disbeliever is only saying: "I do *not* believe there is this negatively defined thing. I do not believe the search will be successful. I have faith that all experiments with positive results could be successfully debunked." So the disbeliever is in a kind of mirror-image of the believer's position. But of course one can never debunk all the experiments, and there will always be more in the future. So the search is equally open-ended. And the disbeliever has to take notice of those positive results. I am thinking particularly of the results of Carl Sargent, Charles Honorton, Helmut Schmidt, and Robert Jahn. I suggest that if we think these can easily be dismissed then we are only deluding ourselves. One cannot offer simplistic counterexplanations and throw all these results away. I am not saying that these results may not, in the future, succumb to some normal explanation; they may well do so. But at the moment we do not have such an explanation.

Whether you are a believer or a disbeliever you will suffer mental conflict and anguish. So what is the solution? Easy, isn't it? Have an open mind. But human beings are not built to have open minds. If they try to have open minds they experience *cognitive dissonance.* Leon Festinger (1957) first used this term. He argued that people strive to make their beliefs and actions consistent and when there is inconsistency they experience this unpleasant state of "cognitive dissonance," and they then use lots of ploys to reduce it. I

have to admit I have become rather familiar with some of them.

First there is premature closure. You can just pick one theory and stick to it against all odds. But I could not do that after all those years. What I could do was only slightly more subtle; that is, I could prefer one theory and ignore the evidence that goes against it. In this way the believer can dismiss negative results by using all the old arguments: The time, the place, the emotional state, or the "vibes" weren't right. Or the disbeliever can refuse to look at the positive results. You may think I wouldn't refuse, but I have to admit that when the *Journal of Parapsychology* arrives with reports of Helmut Schmidt's positive findings I begin to feel uncomfortable and am quite apt to put it away "to read tomorrow."

Alternatively one can jump on a simple counterexplanation, such as "It's all fraud and delusion." Well, maybe it is, but that too creates dissonance of its own. To go around thinking that everyone is cheating, or deluding themselves, can turn one into a permanently suspicious and miserable sort of person, and it can damage one's self-esteem. Suspecting that some effect is fraudulent and tracking that down systematically is one thing, but approaching everything one hears about as though it must be fraud is destructive.

Then there are other cheap ploys. You can decrease the perceived attractiveness of the opposition. The believer can find it easy to put down one famous critic as a dried-up old professor with no real contact with the field anymore, or a more recent one as having shifty eyes and too bushy a beard! Or the disbeliever can dismiss research on the grounds that all parapsychologists are Scientologists, or are too committed to religious beliefs, or are too dreamy-eyed and vague to be doing "real science." But none of this will really wash. And most of us know it won't. Nevertheless, we go on doing it because it is so very hard to have an open mind.

I have said rather a lot about what negative results do not tell us, but is there anything they do tell us? I think we are now in a position to see that there is. I suggest that, wherever you start in parapsychology, if you base your research on the psi hypothesis then you will be forced to do ever more and more restricted research, to back up into ever less and less testable positions, and to produce ever more feeble and flimsy buttresses to hold your theory together. In the end, whatever the questions you started with, you are forced to ask more and more boring questions until there is only one question left: Does psi exist? That question, I submit, is unanswerable.

This process is not restricted to those who get negative results. Helmut Schmidt is among the best researchers in parapsychology, and he has been forced to ask the question "Does psi exist?" Charles Honorton is another example. He is working on fraud-proof, fully automated procedures, even though he might prefer, as do most people in parapsychology, to do process-oriented research, as I did when I started with my question "Is ESP like memory?"

I think that is the problem with parapsychology, and it is a problem that starts from the very hypothesis of psi. The structure and definitions of para-

psychology are to blame. The negative definition of psi, the hundred years of bolstering failing theories, and the powerful will to find something are at fault. They not only force us to ask, "Does psi exist?" but force us to answer in terms of belief. Where there is no rational and convincing answer, belief takes over, and that is why there are two sides, and such misunderstanding.

"The whole history of parapsychology looks like a string of wrong questions."

Here, it seems to me, lies the crux. All those negative results teach us only one thing, that we have been asking the wrong question. And the whole history of parapsychology looks like a string of wrong questions. Parapsychology is, if it is based on the psi hypothesis, a magnificent failure; not because psi doesn't exist, but because it asks unanswerable questions.

An entirely different aspect of my research was prompted by my personal out-of-body experience. I never entirely forgot it. I went on wanting to understand it and eventually tried to tackle it directly.

The first question I asked was the obvious one: "Does anything leave the body in an OBE?" This question may seem close to the unanswerable "Does psi exist?" but I think it is different enough, or perhaps I was just more ruthless in trying to answer it. From experiments of my own, and from reading the literature, I concluded that we do have an answer. And it is "No." You may have heard about an isolated incident of an OBE when someone correctly read a five-digit number (Tart 1968), or when a cat responded to its owner's out-of-body presence (Morris et al. 1978), but I prefer to look at the whole body of evidence (see Blackmore 1982). I concluded that these were unreplicable and that in general we have enough evidence to answer that there is no real evidence for psi in OBEs, there is no evidence of anything leaving the body, and there is no evidence of effects caused by out-of-body persons.

The next question I asked was "Why does the OBE seem so real?" To someone who has not experienced an OBE this might seem a silly starting point, but those of you who have will probably understand why I asked it. That then set me to ask, "Why does anything seem real?" Here I provided myself an answer that seemed to account for the OBE (Blackmore 1984).

Very briefly, I argued that the cognitive system cannot make its decision about what is "real" or "out there" at the low level of chunks of input. Rather, it makes its decisions at the higher level of global models of the world. That is, it constructs models of the world, and chooses one, and only one, as representing "the world out there."

I next had to ask, "Can this decision go wrong?" And the answer is obviously "Yes." When there is inadequate input—damage to the system, drugs, trauma, or any of the many things that can precipitate OBEs—then it might. But what would happen if it goes wrong, the system loses contact with

reality? I would say that a sensible strategy would be to try to replace the lost input model with the next best approximation—one built from memory. And we know a lot about memory models. For example, as Ronald Siegel (1977) has pointed out, they are often built on a bird's-eye view. We know they are schematized, simplified, and often plain wrong. Indeed, they are just like the OBE world.

I proposed that the OBE comes about very simply when the system loses input control and replaces its normal "model of reality" with one constructed from memory. It seems real because it is the best model the system has at the time, and it is therefore chosen to represent "out there."

This answered a lot of questions about the OBE; especially about the phenomenology of the experience. It also led to some predictions I have successfully tested. For example, if the OBE occurs when the normal model of reality is replaced by a bird's-eye view constructed from memory, then the people who have OBEs should be better able to use such views in memory and in imagery. In several experiments I found that OBEers were better at switching viewpoints, were especially good at imagining scenes from a position above their heads, and were more likely to recall dreams in a bird's-eye perspective. I actually had some positive results at last (Blackmore 1986a)!

This theory also led to a new approach to altered states of consciousness in general. To that persistent question "What is altered in an altered state of consciousness?" I could now answer that a person's "model of reality" is altered. I could look at changes induced by meditation, drugs, hypnosis, or a mystical experience, in terms of the changing models of reality (Blackmore 1986b). The OBE could then be seen as only one of a variety of experiences that become possible when the input-driven model of reality is lost.

Interestingly, this theory treats the OBE as a kind of error of reality modeling. And so once again the error can be used to throw light on the normal process at work. But I was only able to come back to this insight once I had abandoned looking for psi. It wasn't that I had rejected the possibility of psi, I had simply ignored it.

"I propose that the OBE comes about when the cognitive system loses input control and replaces its normal 'mode of reality' with one constructed from memory."

I mention my OBE research only to contrast it with my previous work based on psi. In my early work, starting from the psi hypothesis, I was forced to ask, "Does psi exist?" In this research I never had to ask it. The other difference is that I no longer had to worry about having an open mind. That makes me wonder what it is like in other sciences. Of course it is always important to have a potentially open mind. If one's results show that one's hypothesis is wrong, then one has to be prepared to change it; but that need

not happen very often—at least if one's hypotheses are any good it shouldn't. One doesn't have to have a permanent open mind. And so it was with the OBE research—and what a relief!

I can conclude that all my negative results did teach me something. Or am I perhaps only trying to get my 50-cents worth? A few years ago I read an article in the *British Psychological Society Bulletin* about the "Royal Nonesuch of Parapsychology." The author, H. B. Gibson (1979), described Mark Twain's wonderful story of cognitive dissonance, about the show that never was. Many people were lured into paying 50 cents to see a nonexistent show, but instead of decrying the fraud they went out and persuaded others to see it and pay their 50 cents too. Gibson was reminded of this tale, he said, by a conference paper given by a woman who had spent two years in fruitless research on parapsychology. He suggested that parapsychology is only kept going by the "very human tendency to try to get one's 50-cents worth after one has been misled . . . by an unkind fate which has led one into an immense expense of effort in a blind alley."

I fought back in print (Blackmore 1979), arguing that I was not just trying to get my 50-cents worth, that I was after the truth and an understanding of the Nature of Life, the Universe, and Everything. But the problem is that it is very hard to understand the nature of life, the universe, and everything, if you start with the psi hypothesis.

In the end I think my negative results told me that the psi hypothesis leads only to unrepeatability (Blackmore 1985). It forces us to ask ever more boring questions, culminating in the question "Does psi exist?" and to that question there is no obviously right answer. Where there is no right answer, we are in ignorance; and, where we are in ignorance, we should do only one thing—have an open mind. But that is too difficult. After all these years of research, I can only conclude that I don't know which is more elusive—psi or an open mind.

References

Blackmore, S. J. 1979. Correspondence. *Bulletin of the British Psychological Society,* 32:225.
———. 1980a. Correlations Between ESP and Memory. *European Journal of Parapsychology,* 3:127-147.
———. 1980b. A Study of Memory and ESP in Young Children. *Journal of the Society for Psychical Research,* 50:501-520.
———. 1981a. The Effect of Variations in Target Material on ESP and Memory. *Research Letter* (Parapsychology Laboratory, Utrecht); 11:1-26.
———. 1981b. Errors and Confusions in ESP. *European Journal of Parapsychology,* 4:49-70.
———. 1982. *Beyond the Body.* London: Heinemann.
———. 1983. Divination with Tarot Cards: An Empirical Study. *Journal of the Society for Psychical Research,* 52:97-101.
———. 1984. A Psychological Theory of the Out-of-Body Experience. *Journal of Parapsychology,* 48:201-218.
———. 1985. Unrepeatability: Parapsychology's Only Finding. In *The Repeatability Problem in Parapsychology,* edited by B. Shapin and L. Coly, 183-206. New York: Parapsychology Foundation.
———. 1986a. Where Am I? Viewpoints in Imagery and the Out-of-Body Experience. *Journal of Mental Imagery* (in press).

———. 1986b. Who Am I? Changing Models of Reality in Meditation. In *Beyond Therapy*, edited by G. Claxton, 71-85. London: Wisdom.

Blackmore, S. J., and Troscianko, T. S. 1985. Belief in the Paranormal: Probability Judgements, Illusory Control and the "Chance Baseline Shift," *British Journal of Psychology*, 76:459-468.

Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Evanston: Row Press.

Gibson, H. B. 1979. The "Royal Nonesuch" of Parapsychology. *Bulletin of the British Psychological Society*, 32:65-67.

Honorton, C. 1977. Psi and Internal Attention States. In *Handbook of Parapsychology*, edited by B. B. Wolman, 435-472. New York: Van Nostrand Reinhold.

Morris, R. L., Harary, S. B., Janis, J., Hartwell, J., and Roll, W. G. 1978. Studies of Communication During Out-of-Body Experiences. *Journal of the American Society for Psychical Research*, 72:1-22.

Sargent, C. L. 1980. Exploring Psi in the Ganzfeld. *Parapsychological Monographs* No. 17. New York: Parapsychology Foundation.

Siegel, R. K. 1977. Hallucinations. *Scientific American*, 237:132-140.

Spinelli, E. 1983. Paranormal Cognition: Its Summary and Implications. *Parapsychology Review*, 14(5):5-8.

Tart, C. T. 1968. A Psychological Study of Out-of-the-Body Experiences in a Selected Subject. *Journal of the American Society for Psychical Research*, 62:3-27.

[26]

BOUNDLESS MIND: COMING OF AGE IN PARAPSYCHOLOGY

By Marilyn Schlitz

We are alive at a remarkable time in human history. Never before have so many different worldviews, belief systems, and cultures come into contact, interfacing with each other. Just yesterday I was walking through the park a block away from this hotel in Freiburg, Germany. I observed a Chinese man doing his Qigong exercises in one corner. As I walked with my son a little further, I saw a Muslim couple praying; the man stood looking east toward Mecca while the woman stood toward the west. It was a stylized production of something deep and culturally symbolic—and clearly distinct from the Qigong practitioner moving by himself on the lawn a little further away. It was also distinct from the young German boy who was playing an American Indian as he raced past a Sikh man dressed in his turban.

In the past, when people of difference came together, the world was a big enough place that they could diverge to the furthest reaches of the planet. Today the world appears smaller and smaller, and our opportunities to escape that which is different from ourselves are becoming more and more limited. Instead, we find that people with different ontologies and epistemologies are having to come to terms with each other. As a civilization we are having to deal with the fact that life is more complicated and that the nature of reality is much more multifaceted than anything we had previously understood from the narrow focus of our own cultural filters.

As I think now about the convergence of worldviews, I am aware of the kinds of successes we have had as a result of our own Western model of reality. We have cloned a sheep named Dolly and invented a computerized chess champion named Deep Blue. We now have access to the world's wisdom traditions through the World Wide Web. What is happening as a result of all of these rich opportunities? On the one hand, we find ourselves in a period of confusion and conflict as people have to face contradictions and seemingly irreconcilable differences. On the other hand, we also face opportunities for the birthing of new forms of human experience. As we are poised at the beginning of the 21st century, it is an opportunity for parapsychologists to play

---

This article was the Presidential Address delivered at the 43rd Annual Convention of the Parapsychological Association in Freiburg, Germany, August 17–20, 2000.

a fundamental role in speaking to this global situation and to this historical moment. It is a time of opportunity for us.

When I think about the human genome project, for example, I think about the remarkable success of the materialist paradigm—of the physicalist paradigm—that allows us to create a book that maps the human genome. The implications and possibilities from that mapping leave many, many things ahead for us in terms of our ability to diagnose and treat disease, as well as to make changes in future life forms.

But there are many questions left unanswered through an exclusively physicalist model of reality. As we have cultivated this remarkable set of knowledge-based and reason-based skills, there are questions we are not answering. What does it mean to be human? What does it mean to have emotion? What does it mean to have motivation, intention, and attention? All of these first-person experiences that we think of as uniquely our own are not addressed within the strictly physical dimensions of reality. In parapsychology, although we often do not think about ourselves as metaphysicians, we are in a position to arbitrate between these two dimensions of experience: the physical dimensions of reality and the metaphysical dimensions of experience.

Learning our ABCs

As I frame my talk tonight, I recall the delightful Presidential Address by Dean Radin last year. Dean used "Green Eggs and Ham" as his metaphor for describing the state of the field (Radin, 2000). Following his lead, I suggest that we go back to thinking about our place in terms of the basics—our ABCs.

A stands for Action. I believe it is time for us to own our social responsibility as participants in this evolving story of human complexity. I believe we have something to do in transforming the world, such that it becomes a more holistic, integral, and life-affirming scenario for future generations.

B stands for Boldness. Here I would ask all of us to think about the ways in which we have been beaten down by our interest in parapsychology. Have you ever felt you needed to apologize for your interest, particularly when surrounded by mainstream scientists? It is time for us to acknowledge that we are addressing some of the fundamental issues of our time or anytime in human history.

C stands for Context. And that is "storying." It behooves us to see our data as relevant to larger social and political issues. We must answer the "so what" question by touching people at the level of what is important to them and their lives. This leads me to the context for my own participation in psi research.

Coming of Age in Parapsychology

My experience in this field has given me the opportunity to work with the best minds across a variety of testing paradigms. Approaching this

from a first-person perspective allows me to identify some of the ideas that led me to conclude that psi research is relevant and timely.

My story begins in Detroit, Michigan, where I was born and raised. In many ways Detroit is the epitome of the materialist paradigm, a place where the mechanical worldview was perfected. But as a child I was always intrigued by the mystery—I think all children are. And children hear about it as something that, even as adults, we should not fear. I found that as my interest in the mystery developed, I was in good company. It was Albert Einstein who said, "The most beautiful experience we can have is the mysterious...the fundamental emotion which stands at the cradle of true art and true science" (Einstein, 1954, p. 11).

Detroit was a city in crisis during the 1960s and 1970s. During the race riots of 1967, for example, my father was almost killed. The public schools were like a war zone. Fortunately, my parents owned a boat on Lake Sinclair, and this gave me ample time on weekends to spend in nature and away from the harsh realities of Detroit. During that time, I had the chance to develop magical thinking rather than the defeatist attitude that many of my peers developed.

There was something about the mysteries of consciousness that drew me, even as a child. My mother sometimes talked of her experiences with table-tilting. Then, as I was moving into junior high and high school, I had the opportunity to sit next to a guy who was interested in astral projection, and I thought "Wow! This is really far out! How interesting that somehow consciousness might be able to leave the body." And coming from the Sixties and the *zeitgeist* of rebellion, it suddenly seemed to me there were other things beside the materialist model that may be important and may be powerful for understanding life.

It was a time when it was seen as our responsibility to question and challenge authority, question and challenge assumptions that were seen to limit our capacity to break out of a mold the industrialized worldview had created for us. It was a time when our culture yearned for something grander.

As an undergraduate at Wayne State University, I discovered Thomas Kuhn. The now famous social historian wrote about a victory of one paradigm over another involving a matter of faith in a given set of assumptions. Coming out of the Sixties and this rebellious impulse, there was a notion that somehow our reality and that thing called "scientific truth" were were not absolute or fixed. At Monteith College, Wayne State University, we took the curriculum articulated by Kuhn in his book, *The Structure of Scientific Revolutions* (T. S. Kuhn, 1970). We read the classics in the area of paradigm shifts, revolutions in thought, and changes in worldview. This was an inspiration to me, an opportunity to ground what seemed like a fairly groundless form of rebellion into something that had academic merit and intellectual satisfaction.

During this time, I began a mentorship with a psychophysiologist at Wayne State Medical School. We used to talk about the mysteries of the mind and the nature of human consciousness. Before I met this teacher, I had been focusing on the revolution that happened in geology about a

hundred years ago, the area of plate tectonics. Suddenly I was encouraged to think about consciousness and the fact that there are many more mysteries about human experience than the best of Western science has been able to define. I found myself in the middle of something that seemed more current and more timely than plate tectonics.

About this time I picked up a book by the astronaut Edgar Mitchell, *Psychic Exploration* (Mitchell, 1974). This book literally changed my life. I read it and became fascinated by the fact that there were serious scientists taking a serious look at the paranormal. Here was an opportunity to begin to think about a paradigm, in Thomas Kuhn's terms, that was not a hundred years ago but was current and alive. I then switched from geology to psychology.

I later came across the book *Mind-Reach* by Targ and Puthoff (1977). Here the authors claimed that everybody has ESP and that everybody can do these remarkable things like describe a circular building with a white dome at a great distance. There were many examples of this in their book and many claims within the book about the possibility that every person has these potentials.

This led me to begin exploring in the area of remote viewing. That summer I worked with a psychologist, Charlie Solley, doing informal experiments together. I learned a couple of things during this period of discovery. One happened when we brought in a "psychic" to be in one of our experiments. First, I served as the outbound person, and the psychic stayed in. When I came back, she described very little that matched with the description of the site where I was visiting, but Charlie named it pretty well. Then it was my turn to be the inbound investigator. This was the first time I had ever worked with a "subject." I was determined to be a good scientist. That involved objectivity and detachment and the kinds of things that we value as good scientists within the kind of physicalist, materialist model. The woman described a lot of things and I took careful notes. Charlie came back and listened to her description. He then said to me, "What were your impressions?" I said, "I didn't have any." After all, I was the experimenter. And he said "Well, you must have gotten something." I said, "Okay," and made a quick drawing. He got very excited. We went over to the building that served as the target site. The building was surrounded by a moat, the moat was surrounded by a fence, the fence was made up of these symbols, and etched in concrete on the side of the building was the symbol I had drawn. It was early in my exploration that I had to question the assumptions about objectivity and the way in which experimenters can remove themselves from that process of inquiry.

I was so excited about the field of parapsychology that I wrote to the American Society for Psychical Research. They had a directory of educational opportunities in the field of parapsychology. I had friends in southern California and learned that Robert Morris was teaching a course at the University of California, Irvine. I made my way to his class and had the opportunity to do my first formal remote viewing experiment under Bob's tutelage (Schlitz & Deacon,1979). My goal was to test the hypothesis that unselected

people have psi abilities. I was one of the people whom Bob Morris inspired and taught about the nature of critical thinking and about being careful and rigorous about the questions we ask and the methods we use.

My main motivation in studying with Bob was to get to the Institute for Parapsychology for their Summer Study Program. I was certain that here I would get the skills I needed to become a revolutionary in a paradigm shift such as the ones Thomas Kuhn talked about. Having been accepted after applying twice, I had a wonderful time over that summer; anyone who has been through that program knows what an exceptional and life-changing experience it can be. It was for me—as were the next 3 years that I worked there as a research fellow.

I was the last person hired while J.B. and Louisa Rhine were alive. Both were people who represent the kind of bridgework we are called to in parapsychology. J.B. had been a divinity student who then studied biology. He was interested in exploring the interface between that transcendent or extended aspect of human experience and then grounding it within the context of rigorous scientific research. Louisa Rhine's interest in the mysteries of consciousness took the form of spontaneous case reports, again acting as a bridge person between that first-person experience of psi and the scientific laboratory work she inspired.

As we all know, their work was in the area of the forced-choice card-guessing experiments. I was much more interested in the free-response work. I believe I was the first person to do a formal free-response experiment at the Institute for Parapsychology. This was an opportunity for us to begin to think about moving beyond the limits of the restricted range of target response possibilities that we find in the forced-choice experiments and to begin to think about methodologies that might more closely match real-world experiences.

I did a replication of the remote viewing experiment with unselected subjects that I had done under Bob Morris's guidance (Schlitz, 1981). We found chance results, although in the predicted direction. This did not dissuade me. I turned my attention back to the role of the experimenter. Maybe there was something about the fact that we chose this as a career path that would make us good subjects.

About this time I met Elmer Gruber. Together we did the transcontinental remote viewing experiment that had the largest effect size in the remote viewing literature (Schlitz & Gruber, 1980). Both Hans Bender and K. Ramakrishna Rao were the people to whom we sent our data; they were the crossover people to help control for fraud. This again led me to think that experimenters might be a rich source of psi—leading me further from the study of unselected people.

I became interested in the area of healing research. While at the Institute for Parapsychology, I conducted a replication of the Watkins and Watkins study involving the resuscitation of anesthetized mice (Schlitz, 1982). After matching pairs of mice, we set them on adjacent photocells with video cameras above each mouse. A random decision determined which video camera would be activated. The goal was to see if a person in a distant

room could accelerate the resuscitation rate of the selected mouse on the video monitor, as compared with the control mouse. I had the pleasure of working with Jim Davis, Jim Kennedy, Jo-Marie Haight, Kanthamani and Rao, Debbi Weiner, Jim Carpenter, John Hartwell, Richard Broughton, and others who walked the halls of the Institute in the early 1980s.

In one formal study, K. Ramakrishna Rao and I did an experimenter competition study (Rao, Sargent, & Schlitz, 1983). Once again, it looked like the experimenters were stronger as sources in these kinds of experiments than any of the subjects we worked with in the pilot studies.

The interest in psi and healing led me in 1982 to move from Durham, North Carolina, to San Antonio, Texas. Working at the Mind Science Foundation with William Braud, who pioneered the area of allobiofeedback or biological psychokinesis, we developed an experimental model for studying psychic healing in the laboratory. Over a series of many studies, we again explored the difference between selected and unselected subjects, finding the experimenters to be some of our best performers (Braud & Schlitz, 1983; Braud, Schlitz, Collins, & Klitch, 1984; Schlitz & Braud, 1985). Now referred to as DMILS, for direct mental interactions between living systems, the work was important for helping bring parapsychology into rapprochement with socially relevant issues. As Michael Creighton told us in his dinner address to the Parapsychological Association, maybe a decade ago, "You will never be accepted until you do things that are relevant, until people say 'I get it, I get why this is important.'"

As I engaged in this research, I became increasingly aware of the rich nature of qualitative experience. With our nearly exclusive focus on statistical outcomes, parapsychologists sometimes miss all the messy stuff that is the richness of human experience. I felt like we were missing something. But how do we maintain the rigor while opening ourselves up to broader possibilities? With these thoughts, I turned to anthropology, where for over a hundred years researchers have been mapping experience and the sharing of consciousness. This led to a PhD in medical anthropology at the University of Texas, Austin. More important, it started me on a wonderful journey working with healers from a variety of different cultures. From such people I have learned much about the metaphysical dimensions of how healing operates.

Around 1986 I had the great delight of working with Chuck Honorton. Together we did a ganzfeld study among music, dance, and drama students at the Juilliard School of the Performing Arts (Schlitz & Honorton, 1992). As it turns out, this was the last experiment conducted at the Psychophysical Research Laboratory, the last formal experiment that Chuck ever did, and the strongest effect size in the ganzfeld literature. This study was driven by the hunch that creative populations who engaged in the training of attention and intention might produce the strongest results. It seemed to be the case. We found a similar result in a study I later did with Helmut Schmidt (H. Schmidt & Schlitz, 1989), in which our highest scorers were people who had meditation or martial arts training. This is an area I hope to pursue in the future.

In 1992 the Mind Science Foundation decided to discontinue their in-house research program. This left me in this wonderfully challenging situation. I had not finished my PhD, I was just out of a job, and my relationship was on the skids. I remember Dick Bierman asking me in the August PA conference in 1992: "Well, what are you going to do with your life?" I did not know. And what I said at the time was: "Whatever it is, I hope it's something I can't perceive from this vantage point."

In fact, within the week, Ed May offered me a job at the Cognitive Sciences Laboratory at Science Applications International Corporation. This was the government-sponsored CIA-DIA (Central Intelligence Agency–Defense Intelligence Agency) program, which has now been largely declassified. It gave me the opportunity to see some of the data that had been collected by the government and to conduct two remote-staring experiments with Stephen LaBerge (Schlitz & LaBerge, 1994, 1997). In these experiments we put a video camera on one person, communicated their image via closed-circuit television, and then instructed a healer to influence the distant person's physiology. In these studies, we found an increase in autonomic nervous system activity during the staring as compared with the control conditions.

Later I worked with Richard Wiseman, who had done several remote-staring experiments and reported chance results. We decided to conduct a formal experiment together (Wiseman & Schlitz, 1997). In this experiment we did everything identically: the same randomization procedure, the same equipment, the same laboratory, the same subject population. Everything was identical except that I worked with half of the subjects and Richard worked with the other half. We both replicated our original findings. This said to me, again, that there was something important about the intentionality of the experimenter and something wrong with the whole notion of objectivity and detachment.

At the same time that I worked with Ed May and Stephen LaBerge, I also held a postdoctoral fellowship in the psychology department at Stanford University. My PhD research had been about the discourse of power relations in the context of healing. I thought about taking the methodology that I developed in my dissertation research. I proposed to do a discourse analysis of controversial science, looking specifically at the skeptic/proponent debate in parapsychology. By examining the nature of the discourse and the claims to rationality that both sides of the debate make, it seems parapsychologists are much closer than we might think to what is considered to be the opposition. I found, in fact, that if you look at the kinds of claims people are making, the self-proclaimed skeptics will say that their purpose is to help cultivate truth and to dispel people of false beliefs (Schlitz, 1993a, 1993b, 1994). And if you then look at what the parapsychologists are saying, they say "We're here to help people understand the nature of reality. We're out to seek truth." My point is that we are skeptics in the Greek sense meaning open-minded. It was a term to be contrasted with dogmatist. Skeptics from this frame are people who challenge assumptions and try to develop a more coherent model of the

nature of reality. To be skeptical in this way is a profound advantage, especially in a field like ours.

Evaluating the Current Scene

So, where am I now? My work is striving toward an integral model. Like we find in the writings of Ken Wilber, who was influenced by Sri Aurobindo: the notion is that we must embrace the first-, second-, and third-person perspectives to understand the fullness of consciousness. The first person is my interiority, my subjectivity, that nature of experience that makes me uniquely me; the second person tells me that I am not uniquely me, that I am embedded in a social context that often shapes the things that I think are uniquely me; and the third person is that aspect of experience which is the "it"—the data that we all talk about.

The analysis I did at Stanford on the skeptic/proponent debate found that while people use "it" language to justify their beliefs, it really comes down to "I" (Schlitz, 1993a, 1993b, 1994). It may be that I had an experience, or that I know somebody whom I trust who had an experience. But it is important that we recognize that even the "it" and the "I" are grounded in the "we," and that we share a certain worldview that structures our perceptions. I now find myself working with a model that integrates the various perspectives.

Today I direct the research program at the Institute of Noetic Sciences. We have about 40 projects on various aspects of consciousness. A number of them fall into the area of parapsychology, although I almost never use that word. It is not because the people in my orbit do not like parapsychology but because they do not think this is where the action is. I have found it valuable to restory what we do so that we can create relevant contexts for sharing our work and for getting the funding that we need.

Noetic is an ancient Greek word to describe direct experience. William James (James & Perry, 1912) used it to describe states of insight unplumbed by the discursive intellect involving illuminations and revelations that, while inarticulate, are full of significance and importance. As a rule, James noted, they carry with them a peculiar sense of authority.

I was recruited to the Institute of Noetic Sciences in 1993. It felt like coming full circle to the organization started by Edgar Mitchell, who wrote the book that started me on this whole expedition. It has been in many ways like coming home.

Onward to the Future

How does all this relate to the issue of parapsychology and where we are going as a field? Clearly, there are many possible futures. There are also many possible ways to evaluate our success.

Participation in the Field of Parapsychology

If we define the field as members of the Parapsychological Association (PA), we have not grown much throughout our history. However, there is a tremendous amount of action happening outside of the PA by people whom we had never heard of and who had never heard of us. In particular, I would say that there are many people now doing research in mainstream medical centers on distant healing, for example, and who are now publishing in mainstream medical journals. Randolph Byrd, for example, conducted the first clinical distant healing study (Byrd, 1988). He studied patients in a cardiology unit, randomizing them to a distant prayer group and a standard care group. He found that the people who received distant prayer or intention recovered better and had fewer medical complications than the standard care group alone. This project was picked out and replicated by a team under the direction of W. S. Harris, who again found that the group who received distant intentionality did better than the controls (Harris et al., 1999).

Elisabeth Targ has just published a study in the *Western Journal of Medicine* (Sicher, Targ, Moore, & Smith, 1998) looking at the effect of distant healing on patients with HIV. She did two studies, both yielding significant results for distant healing. Again, the group who received the distant intentionality seemed to have better outcomes on a variety of different health outcome measures.

I had the pleasure of convening several meetings in which we invited scientists from mainstream medical centers who were interested in or actually doing clinical outcome studies involving distant intentionality. We had about 50 people at the first meeting, representing some 20 medical centers. At the second meeting, we had over 100 people. This is not something that is only happening here within our ranks but has moved very much into the mainstream community.

Global Outreach

In terms of global community, it seems to me that there is reason to feel optimism about the hot spots across the world. Right at this moment, there is a study looking at distant Qigong on cell growth in Beijing, China, by a molecular geneticist named Garret Yount. There are studies going on in South America, Africa, Indonesia, India, and in almost every country in Europe. There is reason to feel optimistic that our kind of work is catching on throughout the globe.

The Internet also provides an important indicator of global community. I look at the PDL, and several other discussion groups that are going on, linking people in a variety of different countries and time zones in an ongoing dialogue. Roger Nelson's project creating a global "EEG" study is a very exciting and novel approach to experimental research (Nelson, 2001). It is also a way of developing community and some kind of fraternity across the globe in these various settings.

Quality of the Science

How have we improved our ability to capture the kind of phenomena that we are looking for? Parapsychologists have always tried their best to create standards that are exceptionally good and that can withstand the scrutiny of critical or negatively oriented reviewers. But if you look at something like the study that Rupert Sheldrake did comparing the use of randomized double-blind protocols in a variety of different academic fields—biology, chemistry, physics, psychology, and medicine—the group that stood out in terms of their standards and their commitment to what is considered the gold standard in Western science—the use of the randomized double-blind protocol—was parapsychology (Sheldrake, 1994). We stood out as the group who did the best quality of research. And with contributions by new members of our community, such as Stefan Schmidt, we are able to improve the quality of our science (S. Schmidt & Walach, 2000). I think the more we subject ourselves to that kind of scrutiny, the higher the quality will be.

Theory Building

Again, we have not solved the problem. We do not know how to predict psi. At the same time I would say that we have made progress. Walter Von Lucadou, Fotini Pallikari, Dick Bierman, and James Spottiswoode—all of these people within our community who are very able, very capable theoreticians, are attempting to understand the complexities of psi and to develop predictive models that might help us to harness and capture the phenomena we are studying. There are people outside our field to whom we should pay attention. There are people like Stuart Kaufman from the Santa Fe Institute, famous for his work in complexity theory, deeply interested in this topic. I just had a delightful week with Ralph Abraham from the University of California, Santa Cruz; he is somebody who is bright and interested in helping us work on this problem. We have got Walter Freeman, professor emeritus from Berkeley, who in his book, *Societies of Brains: A Study in the Neuroscience of Love and Hate,* mentions parapsychology in the appendix (Freeman, 1995). Henry Stapp, a world-class physicist, is also pursuing this work. And there are others. In terms of theory development, we may not be there yet, but as we build bridges and reach out to people within the mainstream, we increase the likelihood that we will be able to come up with a breakthrough.

Research Centers

Much is happening in different labs around the world. We have had the opportunity to hear some from the Institut für Grenzgebiete der Psychologie und Psychohygiene, which represents an important development here in Germany. Mario Varvoglis is resurrecting the Institut Métapsychique International program in Paris. Dean Radin has moved from the Integral Institute to create the Boundary Institute, funded by people from the dot-com world. These are people in Silicon Valley who

have unlimited resources financially and who are just testing the waters with us. This is an exciting opportunity. There is Dick Bierman's work at Starlab. While he is not totally coming out as a parapsychologist in this work, I think he is on the edges pursuing his psi research and presentiment data wherever he can.

I am building a lab right now in partnership with the California Pacific Medical Center. Elisabeth Targ and I are joint directors of this program. California Pacific Medical Center is a mainstream medical center and one of the oldest hospitals in the United States. On their invitation, we are now building what we hope will be a leading center for distant healing and subtle energies research within the context of a mainstream medical center.

Ruth Reinsel just told me about the Neuroscience Laboratory that she just started. There are other programs that represent mainstream medical centers, programs, people that we have never heard of. There are distant healing research projects going on at Duke, Scripps, Harvard, California Pacific Medical Center, University of California, Irvine, and University of California, San Francisco. It is happening. And it is time for us to pay attention and look at the way in which there is currency and context in these studies, and to show that the application of some of our work is important within the clinical context.

Publications and Conferences

On the negative side, there was a recent survey reported at the American Psychological Association which found that in the mid-1980s reference to parapsychology occurred in about 67% of the psychology textbooks for the introductory level. In the last decade, it has gone down to about 50% (Roig, Icochea, & Cuzzucoli, 1991). The authors report that parapsychology research, once one of the most sought after courses at St. John's, is not even given anymore. Roig and his colleagues argue that interest in the paranormal in the general public has not waned, but interest in parapsychological research has.

On the positive side, there is a new introductory textbook on psychology in the United Kingdom that is doing extremely well. There was the International Congress of Psychology in Stockholm. The University of Arizona at Tucson Consciousness Conference has an ongoing interest in our field, with several parapsychology presentations at the last conference. This conference is a forum for doing bridgework between the kind of behaviorist or physicalist programs in consciousness studies and more transpersonal approaches. From this conference, Ed May and I had the opportunity to publish a chapter on parapsychology that came out in the last year through MIT Press. This is all in the last year (Schlitz & May, 1998).

A program called "Closer to Truth" was started by Robert Kuhn, a lawyer who is interested in the furthest reaches of consciousness. Parapsychology is one of his major areas of interest. It started with a television program creating debates about controversial topics in the area of consciousness. He then moved from creating a 52-part television series to a book that is now published by McGraw-Hill (R. L. Kuhn, 2000). He also

created a Web program in which people ranging from Nobel laureates to neuroscientists to philosophers are debating the data for parapsychology and other anomalies.

Cardeña, Lynn, and Krippner's (2000) book, *Variety of Anomalous Experiences*, which was just published by the American Psychological Association, included a chapter on parapsychology. There is a lot happening in the area of publications and conferences that I cannot even begin to touch.

Education

All of us should take our hats off to Bob Morris. I think he is doing more than anybody else to fuel the field by bringing new, innovative talent through the program at the University of Edinburgh. In the United Kingdom alone, 50 PhDs have been graduated in the area of parapsychology (Carr, 2000). In the United States, we have programs like Saybrook, the California Institute of Integral Studies, the Institute of Transpersonal Psychology—all of them producing new PhDs.

Career Opportunities

Again, we find that 20 of the graduates from the University of Edinburgh have now obtained permanent academic positions. In the United States, I would say that the whole area of alternative medicine is exploding. I now have the opportunity to serve on a congressionally appointed committee at the National Institutes of Health (NIH). I was selected because of my interest in parapsychology, not in spite of it. Because of the groundswell of interest in alternative and complementary medicine, all of the major medical centers in the Unites States are developing programs. They have to do research to complement the clinical applications. They are interested in psi research because we have rigorous methodological programs to look at the interface between the physical and the metaphysical. Members of the complementary and alternative medicine community find this inspiring and encouraging, and there are job possibilities there if you can begin to package and create a context that shows the relevance of what we are doing.

Integration Into the Mainstream

In addition to work at the NIH, in the past year I have talked to Harvard Medical School, Stanford Medical School, and the Mayo Clinic. These are programs that were once completely resistant to this kind of work. Today they are beginning to open the door to us in new ways. I would say that our ability to penetrate the mainstream is really beginning to work.

Funding

The big question: Funding is always the problem for everybody who is trying to do work in a marginal area—again I will say that NIH is a bright star within the United States. Elisabeth Targ has just received two

major grants to do distant healing research. Garret Yount just received a grant from NIH, and Leanna Standish and I received a grant to do a study looking at brainwave synchronization between two people at a distance.

We have got the Parrott-Warrick—stellar advances that are happening there in terms of bringing funding into the field. The Parapsychology Foundation is now offering a new grant. It has been a good friend to this field for a long time, not giving major funding but enough to keep a lot of people, including myself, from going on the dole. The Bial Foundation in Portugal has been an important source for funding in recent years. The Fetzer Institute has put money into a large project on psi. Laurence Rockefeller is now funding a project that we are doing on distant healing.

Conclusion

By way of conclusion, I do want to remind us of where we began—the ABCs stand for Action, Boldness, and Context.

Action

It is important to speak to socially relevant issues. It is important for us to recognize that what we are doing is fundamentally vital. As I talked about converging worldviews and the fact that we are in a time of global transition, I think it is important for us to recognize that we are between stories. The old story is not working anymore, and we can see that. Modern medicine, for example, is in crisis. And at the same time, the new story has not been born yet. We are in this transitional, liminal phase, of waiting to see what are the appropriate questions to be asking about human possibility and about the human condition. And we have a role to play in formulating those questions, if not answering them ultimately, in terms of our ability to make links and to actually resolve some of the questions before us.

Boldness

When I talk about paradigm and I talk about Thomas Kuhn, it all may sound a little old or a little like Pollyanna. It is not. The questions we are asking are really the most important things: What is life? What is consciousness? What is our capacity as human beings to become something more? Is there a new story that is not about a strictly physical, reductionist, separate, objective world out there, but one in which we are fundamental actors in the evolutionary process? If we can be bold enough to own our responsibility, we may be bold enough to recognize that we are conscious participants in an evolving universe.

Context

We need to story our work in a way that is more relevant. We can do a remarkable job of putting anyone to sleep over our findings. We need to wake up to the fact that it is really interesting. If we create the context for

understanding psi research, both in terms of politics and in terms of the social implications, I think the whole field will move forward. Whether we call it parapsychology or distant intentionality or consciousness studies or alternative medicine or applied epistemology—whatever we choose to call it—we are making a difference in the world.

Finally, I want to remind us of where we started, which is in the mystery. We should embrace it and our dance with the ineffable. Our goal is to begin to understand and to play with mystery in a more active way. As I was preparing this talk, my son got sick. I was going through a baby book, and I came across this quote in *The Well Baby Book* (Samuels & Samuels, 1991): "Young children often see and point out to their parents objects or persons that the parents can't see. Such images may be imaginary or real. The existence of real psychic phenomena has been demonstrated in parapsychological experiments over the past 30 years" (p. 224).

The paradigm is shifting, whether we want to go with it or not. I am reminded of Max Planck's infamous quote, "A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die and a new generation grows up that is familiar with it" (Planck & Laue, 1949, pp. 33-34). The paradigm is changing, and it is our job to act as bridge makers to something bigger and something more inclusive of the fullness and the richness of the human condition.

References

Braud, W., & Schlitz, M. (1983). Psychokinetic influence on electrodermal activity. *Journal of Parapsychology,* **47**, 95–119.

Braud, W., Schlitz, M., Collins, J., & Klitch, H. (1984). Further studies of the bio-PK effect: Feedback, blocking; specificity/generality [Abstract]. In R. A. White & J. Solfvin (Eds.), *Research in parapsychology 1984* (pp. 45–48). Metuchen, NJ: Scarecrow Press.

Byrd, R. C. (1988). Positive therapeutic effects of intercessory prayer in a coronary care unit population. *Southern Medical Journal,* **81**, 826–829.

Cardeña, E., Lynn, S. J., & Krippner, S. (Eds.). (2000). *Varieties of anomalous experience: Examining the scientific evidence.* Washington, DC: American Psychological Association.

Carr, B. (2000). *Research activities in the SPR: New initiatives.* Retrieved November 17, 2000 from http://www.spr.ac.uk/research_initiatives.html.

Einstein, A. (1954). *Ideas and opinions.* New York: Crown.

Freeman, W. J. (1995). *Societies of brains: A study in the neuroscience of love and hate.* Hillsdale, NJ: Erlbaum.

HARRIS, W. S., GOWDA, M., KOLB, J. W., STRACHACZ, C. P., VACEK, J. L., JONES, P. G., FORKER, A., O'KEEFE, J. H., & MCCALLISTER, B. D. (1999). A randomized, controlled trial of the effects of remote intercessory prayer on outcomes in patients admitted to the coronary care unit. *Archives of Internal Medicine*, **159**, 2273–2278.

JAMES, W., & PERRY, R. B. (1912). *Essays in radical empiricism.* New York: Longmans Green.

KUHN, R. L. (2000). *Closer to truth: Challenging current belief.* New York: McGraw-Hill.

KUHN, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

MITCHELL, E. D. (1974). *Psychic exploration: A challenge for science.* New York: Putnam's.

NELSON, R. D. (2001). Correlation of global events with REG data: An internet-based, nonlocal anomalies experiment. *Journal of Parapsychology*, **65**, 247–271.

PLANCK, M., & LAUE, M. v. (1949). *Scientific autobiography and other papers; with a memorial address on Max Planck.* New York: Philosophical Library.

RADIN, D. I. (2000). What's ahead? *Journal of Parapsychology*, **64**, 353–364.

RAO, K. R., SARGENT, C., & SCHLITZ, M. (1983). Two series of volitional studies with competition set [Abstract]. In W. G. Roll, J. Beloff, & R. W. White (Eds.), *Research in parapsychology 1982* (pp. 190–193). Metuchen, NJ: Scarecrow Press.

ROIG, M., ICOCHEA, H., & CUZZUCOLI, A. (1991). Coverage of parapsychology in introductory psychology textbooks. *Teaching of Psychology*, **18**, 157–160.

SAMUELS, M., & SAMUELS, N. (1991). *The well baby book: Revised and expanded for the 1990s.* New York: Summit Books.

SCHLITZ, M. (1981). Remote viewing: Observations based on four experimental series [Abstract]. In R. White (Ed.), *Research in parapsychology 1979* (pp.161-162). Metuchen, NJ: Scarecrow Press.

SCHLITZ, M. (1982, April). *PK on living systems: Further studies with anesthetized mice.* Paper presented at the meeting of the Southeastern Regional Parapsychological Association,Durham, NC.

SCHLITZ, M. (1993a, April). *Discourse analysis as a tool for conflict mediation.* Paper presented at the Society for Applied Anthropology Conference, San Antonio, TX.

SCHLITZ, M. (1993b, November). *A discourse-centered approach to scientific controversy: The case of parapsychology.* Paper presented at the meeting of the American Anthropological Association, Smithsonian Institute, Washington, DC.

SCHLITZ, M. (1994, September). *The ganzfeld debate in parapsychology: Truth construction and scientific controversy.* Paper presented at the meeting of the American Anthropological Association, Atlanta, GA.

SCHLITZ, M., & BRAUD, W. (1985). Reiki plus natural healing: An ethnographic and experimental study. *Psi Research,* **4**, 100–123.

SCHLITZ, M., & DEACON, S. (1980). Remote viewing: A conceptual replication of Targ and Puthoff [Abstract]. In R. White (Ed.), *Research in parapsychology 1979* (pp. 124–126). Metuchen, NJ: Scarecrow Press.

SCHLITZ, M. J., & GRUBER, E. (1980). Transcontinental remote viewing. *Journal of Parapsychology,* **44**, 305–317.

SCHLITZ, M. J., & HONORTON, C. (1992). Ganzfeld psi performance within an artistically gifted population. *Journal of the American Society for Psychical Research,* **86**, 83–98.

SCHLITZ, M., & LABERGE, S. (1994). Autonomic detection of remote observation: Two conceptual replications. *Proceedings of Presented Papers: The Parapsychological Association 37th Annual Convention,* 352–360.

SCHLITZ, M. J., & LABERGE, S. (1997). Covert observation increases skin conductance in subjects unaware of when they are being observed: A replication. *Journal of Parapsychology,* **61**, 185–196.

SCHLITZ, M. J., & MAY, E. (1998). Parapsychology: Fact or fiction? Replicable evidence for unusual consciousness effects. In S. R. Hameroff, A. W. Kaszniak, & A. C. Scott (Eds.), *Toward a science of consciousness: II. The second Tucson discussions and debates* (pp. 691–700). Cambridge, MA: MIT Press.

SCHMIDT, H., & SCHLITZ, M. J. (1989). A large-scale pilot PK experiment with prerecorded random events [Abstract]. In L. Henkel & R. Berger (Eds.), *Research in parapsychology 1988* (pp. 6–10). Metuchen, NJ: Scarecrow Press.

SCHMIDT, S., & WALACH, H. (2000). Electrodermal activity (EDA): State-of-the-art measurements and techniques for parapsychological purposes. *Journal of Parapsychology,* **60**, 139–164.

SHELDRAKE, R. (1994). *Seven experiments that could change the world.* London: Fourth Estate.

SICHER, F., TARG, E., MOORE, D., & SMITH, H. S. (1998). A randomized double-blind study of the effect of distant healing in a population with advanced AIDS: Report of a small scale study. *Western Journal of Medicine,* **169**, 356–363.

TARG, R., & PUTHOFF, H. (1977). *Mind-reach: Scientists look at psychic abilities.* New York: Delacorte.

WISEMAN, R., & SCHLITZ, M. (1997). Experimenter effects and the remote detection of staring. *Journal of Parapsychology,* **61**, 197–207.

Institute of Noetic Sciences
101 San Antonio Road
Petaluma, CA 94952-9524, USA
schlitz@noetic.org

Name Index

Abelson, R.P. 317
Abraham, Ralph 486
Ahmed, I. xix
Akers, Charles xvii, 253–315, 317, 326
Alcock, James E. xviii, xix, 253, 287, 288, 324, 325, 326, 441–62
Alexander, Cheryl 366
Allport, G.W. 50
Anderson, C.A. 47
Andrews, Sperry 401–19, 421, 422
Annemann, Theodore 151
Asher, S.J. 49
Astin, John A. xix, 433–40
Atkinson, G. 67, 68
Atkinson, R.C. 330, 357
Atkinson, R.L. 330, 333, 357
Auerbach, Lloyd 100, 149
Austin, V.M. 212, 217
Avant, L.L. 330
Ayeroff, F. 317

Baggally, W.W. 163, 165, 166
Baker, R.A. 65, 66
Banerji, D.K. 176, 179, 181
Banerji, Mrs 176
Banks, I. 141
Barber, T.X. 269, 280
Bar-Hillel, M. 40
Barker, P. 255
Barnier, A.J. 67
Barrington, M.R. xvi
Bartlett, F.C. 194
Bartlett, H.J. 255, 268, 278, 294, 301, 302
Bateman, F. 453
Bayarri, M.J. 341
Beloff, John xiii, xvi, 112, 149, 153–71, 255, 257, 265, 266, 276, 290, 292, 300
Belvedere, E. 322, 323
Bem, Daryl J. xvii, xviii, 329–43, 346, 348, 349, 355, 356, 357, 358, 366
Benassi, V.A. 151
Bender, Hans 140, 143, 144, 149, 173, 201, 206, 212, 213, 214, 215, 217, 219, 481
Benor, D. 434
Bent, D.H. 8
Berdsani. P. 149
Berger, J. 341
Bergson, Henri 162
Bessent, Malcolm 217, 227
Besterman, T. 58
Beuno, St 132
Beutler, J.J. 437
Bevan, J.M. 255, 257, 258, 267, 278
Beyerstein, D. xvi
Beyerstein, B.L. 457, 458
Bhadra, B.H. 255, 286, 289, 301
Bhagawantam, S. 181
Bhattacharya, P.K. 181
Bierman, Dick J. 263, 365, 366, 483, 486, 487
Billings, F.J. 67
Birge, W.R. 211, 215, 216, 217, 223
Bjork, R.A. 334
Björnsson, Hafstein 218
Blackmore, Susan J. xvii, xix, 41, 64, 65, 68, 70, 317, 458, 465–76
Blatavsky, Mme 161
Blondlot, Rene 447
Bloxham, Arnell 66
Boerenkamp, H.G. 220, 221, 224, 226
Borgida, E. 40
Boring, E.G. 287
Bosga, Douwe 219, 366
Botazzi, Philippe 162
Bozarth, J.D. 331
Bradley, James V. 326
Bransford, J.D. 48
Braud, L.G. 255, 258, 259, 265, 268, 274, 277, 278, 279, 288, 290, 294, 295, 330
Braud, William G. xviii, xix, xx, 51, 255, 258, 259, 265, 266, 268, 274, 277, 278, 279, 288, 290, 292, 294, 295, 301, 322, 330, 366, 401–19, 421, 422, 437, 482
Brekke, N. 40
Bridgman, P.W. 287
Briggs, K.C. 408
Broad, W.J. 276
Brogan, D. 100
Broughton, Richard S. xviii, 263, 282, 283, 330, 338, 340, 357, 482
Brown, G.S. 287
Brown, L.S. 62
Brown, R. 60

Brugger, Peter 459, 460
Bruno, J. 151
Buck, G.E. 243
Burdick, D.S. 216, 223, 288, 289
Bursik, K. 69
Bush, R.R. 321
Butler, D.C. 253, 287
Byrd, Randolph C. 435, 438, 485
Bzdek, V.M. 436, 437

Cadoret, R.J. 254, 255, 257, 265, 266, 267, 279
Calkins, J.L. 287
Caradoc, Prince 132
Cardeña, E. xv, 488
Carlson, E.B. 67
Carpenter, James C. 255, 257, 265, 266, 271, 280, 281, 285, 287, 356, 482
Carr, B. 488
Carrington, Hereward 149, 161, 163, 165, 166, 173, 174
Casler, L. 254, 255, 257, 258, 265, 267, 269, 277, 278, 279, 280, 281
Casper, G.W. 255, 257, 258, 275, 278, 279
Ceci, Stephen J. 46
Chang, A. 255
Chapman, J.P. 45
Chapman, L.J. 45
Cherlow, D.D. 102
Child, Irvin L. xvii, 273, 289, 317–28, 330
Christopher 149
Clancy, S.A. 70
Clark, T.K. 253
Cochran, W.G. 263
Cockayne, E.A. 77
Cohen, G. 57
Cohen, J. 333, 434
Cole, P. 291
Coles, M.G. 417
Collins, H. 149
Collins, J. 482
Collipp, P.J. 435, 437
Conway, M.A. 62
Cook, A.M. 68
Cook, E.W. 78
Cook, Florence 167, 174
Coover, J.E. 421, 422
Corinda, T. 151
Cornell, A.D. 99, 112, 140
Corty, Eric 43, 44
Costa, P.T. jr 340
Couer, Jacques 66
Courtier, Jules 162
Cox, G.M. 263
Cox, W.E. 144, 149, 150
Cranall, P.H. 102
Crandall, J.E. 317
Crandon, Margery 275, 276
Creighton, Michael 482
Croiset, Gerard 212–15 *passim*, 216, 218, 219, 223
Crookes, William 159, 174
Crumbaugh, J.C. 297
Crussard 149
Cunningham, S. 357
Curie, Marie 162
Curie, Pierre 162
Cushing, J.T. 341
Cuzzucoli, A. 330, 487

Dabney, W.C. 89
Dale, Laura A. 210, 286
Dalgleish, T. 62
Dalton, Kathy 347, 348, 360, 365, 366
Damgaard, J.A. 217
Daniels, Michael xv, 117–45
Davey 58
Davies, G. 62
Davis, Jim 77, 78, 482
Davis, J.W. 257, 258, 263, 264, 265, 282, 289
Deacon, S. 480
Dean, D. 217
de Broglie, Louis V. de 385
Delanoy, Deborah 366
Deese, J. 62
Denaro, S.J. 77
Dennis, Michael 4
De Sano, C.F. 68
Devant, D. 151
Diaconis, Persi 35, 36, 37, 38, 39, 149, 253, 258, 288, 289
Diamond, M.J. 68
Dingwall, Eric J. 155, 163, 295, 330
Dobson, M. 67
Doll, R. 89
Don, Norman 366
Dooley, Concepta 118, 125–6 *passim*, 141, 142
Dooley, Michael 118, 125–6 *passim*, 141, 142
Dossey, L. 439
Dow, C.A. 44
Dronek, E. 260, 261
Druckman, D. 334
Dukhan, H. 255, 257, 265, 266, 271, 280, 281, 297
Dunne, J.W. 238, 239, 240

Edge, Hoyt 362
Edwards, P. xv
Einstein, Albert 385, 446, 479

Eisen, M.L. 67
Eisenbud, J. 149, 174, 218, 276
Elfferich, Ingrid 91–7
Elliott, Mrs Warren 204
Ellison, A. 444
Eppinger, Ricardo 366
Ernst, Edzard 433–40
Ertel, S. xvii
Erwin, William 321
Esser, A.H. 216
Estabrooks, G.H. 277
Evans, Jane 66
Evans, J. St B.T. 40, 44, 50
Eysenck, Hans J. 340, 417

Fahler, J. 254, 255, 257, 265, 266, 267, 279
Falk, R. 37, 38, 48, 49
Fanibunda, Eruch 182
Faraday, M. xviii, 373–5
Fatteh, A. 82
Feller, W.K. 258
Fielding, Everard 163–6 *passim*
Femina, D.D. 61
Ferrari, D.C. 338, 349, 350, 352, 355, 356
Festinger, Leon 471
Fiedler, K. 42
Finch, S.E. 68
Findlater, J.W. 203
Fischhoff, B. 37, 38, 42, 50
Fiske, S.T. 41, 44
Fitzkee, D. 151
Flammarion, Camille 161
Flournoy, Theodore 161
Flynn, C.P. 141
Foà, Pio 162
Foley 293
Fontana, D. xv, xvi, 444
Forer, B.R. 187
Foulkes, D. 322, 323
Fox, W. 255
Franklin, I. 100, 101, 103
Franks, J.J. 48
Frantz, C.H. 86
Fraser, J.T. 102
Frazier, K. 253
Freedman, D. 422
Freeman, Walter 486
French, Christopher C. xv, 55–76
Friend, R. 407, 416
Fuller, Uriah 150

Gale, A. 417
Galloway, T.M. 89
Gardner, M. 149, 253, 260, 281, 282, 291, 296, 454
Garrett, Eileen 201, 206–7 *passim*
Garry, M. 63
Gatlin, L.L. 260, 261, 287
Gauld, A. 99, 102, 112, 140
Gearhart, L. 102
Geen, R.G. 417
Geley, Gustave 167
Geller, Uri 149, 253, 257, 269, 276, 292, 296
Gengerelli, J.A. 255, 257, 268, 270, 277, 285, 287
Gerber, D.M. 84
Gerding, Hans 366
Gerding, J.L.F. 220
Gibson, E.P. 243
Gibson, H.B. 475
Gibson, N.Y. 243
Gigerenzer, G. 42, 43, 44, 50
Gilovich, T. 329
Girden, E. xviii, 253, 262, 282, 287, 299
Globus, G. 322, 323
Gloye, E.E. 149
Gokak, V.K. 181
Goldberg, J. 218
Goldman, A. 261
Goldney 256
Goodfellow, L.D. 289
Gorbachev, Mikhail 36
Gordon, A. 436, 438
Gordon, I. 82
Gower, David 117, 118, 122, 133, 136, 138, 139, 141, 142, 143
Gower, John-Paul 117, 118, 119, 120, 121, 124, 125, 133, 134, 137–40 *passim*, 141, 143, 144
Gower, Rose-Mary 117, 118, 119–23 *passim*, 124, 125, 130, 131, 132, 134–40 *passim*, 141, 142, 143, 144
Greeley, A.M. 4, 25
Green, C.E. 4, 13, 30, 330
Greening, Emma K. 58, 68, 69, 99–115
Greenwood, J.A. 256, 290
Grela, J.J. 255, 257, 258, 278, 295
Greville, T.N.E. 211, 289
Grey, W. 56
Greyson, B. 69, 95, 437, 438
Groome, D. xv
Groth-Marnat, G. 68, 69
Gruber, Elmer 481
Gudjonsson 70
Gudmundsdottir, A. 4
Guiley, R.E. 100

Haight, Jo-Marie 482
Halgreen, E. 102
Hall, Calvin 322
Hall, D.F. 49

Hall, J. 386
Hall, S. xix
Hannesson, G. 173
Hansel, C.E.M. xviii, 253, 256, 259, 260, 262, 263, 264, 270, 275, 276, 281, 283–4 *passim*, 291, 296, 299, 303, 323–4 *passim*
Hansel, Mark 166
Hansen, George P. xvi, 149–52, 218
Haraldsson, Erlendur xiv, xvi, 4, 69, 70, 149, 150, 173–83, 218, 255, 257, 258, 259, 265, 277, 279, 280, 281, 282, 297, 300
Hardy, A. 287
Harkness, Elaine 433–40
Harley, T.A. 255, 286, 287
Harper, S. 255, 257, 268, 273, 274, 278, 289, 294, 298, 330
Harvie, R. 287
Harribance, Sean 217
Harris, M.J. 66, 332, 334, 338
Harris, W.S. 435, 485
Harsch, N. 60
Hart, H. 13
Harte, T.M. 99
Hartwell, John 482
Harvie, R. 287
Haskell, P. 256, 299
Hasra, Professor 178, 179
Hasted, J. 149
Hastings, Arthur 149
Havens, R.A. 99
Hayes, N. 357
Healey, J. 322
Heaps, C. 67
Hebb, Donald O. 457
Heijn, G.J. 220
Heisenberg, Werner K. 385
Henderson, J. 100
Henning, Douglas 182
Herbert, N. 341
Henry VIII, King 100
Herbert, C.V.C. 207, 264
Hettinger, 208–10 *passim*
Hill, A.B. 89
Hintzman, D.L. 49
Hirdaramani, L. 177
Hite, D.D. 317
Hodgson, Richard 161
Hoebens, P.H. 219
Hogarth, R.M. 50
Holden, K.J. 66
Holinrake, Fran 100, 108
Holmes, D.S. 61
Holmes, L.B. 78
Home, Daniel D. 167, 174, 444
Honorton, Charles xvii, xx, 217, 253, 254, 255, 257, 264, 265, 266, 268, 270, 273, 274, 277, 278, 279, 289, 290, 291, 292, 293, 294, 297, 298, 299, 300, 301, 302, 303, 304, 317, 318, 325, 329–43, 346, 347, 348, 349, 350, 351, 352, 355, 356, 357, 358, 359, 361, 366, 468, 471, 472, 482
Hootkooper, J.M. 263
Hope, Lord 174
Hopkins, B. 65
Horselenberg, R. 67
Houck, R.L. 417
Houdini, Harry 167
Houran, J. xv, 99, 100, 102, 111, 113
Hövelmann, Gerd 366
Howard, Catherine 100
Hoy, David 149
Hubbard, G.S. 263, 283
Hughes, J.P.W. 89
Hull, C.H. 8
Hume, David 56, 154–5 *passim*
Humphrey, B.M. 255, 262, 266, 275, 289
Humphrey, B.S. 263, 282
Humphreys, W.C. 262
Hurkos, Peter 216
Hurst, C. 422
Hyman, I.E. jr 58, 63, 64, 67, 149
Hyman, Ray xiii, xvi, xvii, xviii, 185–97, 253, 254, 257, 260, 264, 272, 273, 276, 293, 294, 296, 317, 331, 332, 333, 334, 336, 338, 346, 347, 349, 350, 351, 352, 354, 357, 358, 361, 447, 454
Hynes, L. 67
Hyslop, James H. 202–3 *passim*

Icochea, H. 330, 487
Irwin, H.J. xv, 68, 69, 99, 317
Iverson, J. 66

Jacobs, D.M. 65
Jahn, Robert G. xviii, 322, 448–9 *passim*, 453, 471
James, William 484
Jarvis, Branwen 132
Jastrow, Joseph 165, 166
Jeffers, Stanley 448–9 *passim*, 456, 460
Johnson, Martin 151, 255, 257, 258, 265, 266, 267, 275, 278, 279, 280, 281, 286, 287, 288, 299
Johnson, M.K. 64
Johnston, Wade M. 47
Jones, E.E. 47
Jones, Warren H. xv, 59, 253, 258, 322, 325–6
Jonsson, S. 4

Joshi, S. 69
Joyce, C.R. 435

Kahneman, Daniel 39, 40, 41, 43, 50, 329, 338, 339
Kallai 49
Kammann, Richard xv, 36, 149, 253, 257, 264, 269, 271, 291, 292, 324
Kanthamani, B.K. 255, 257, 258, 265, 266, 267, 271, 275, 278, 279, 280, 281, 286, 287, 288, 300, 301, 302, 322, 338, 482
Kasturi, N. 174
Katz, A.N. 67
Kaufmann, H. 112
Kaufman, Stuart 486
Kaye, Marvin 151
Kebbell, M.R. 61
Keen, M. 444
Keller, E. 436, 437
Kellogg, C.E. 287
Kelly, E.F. 256, 287, 288, 289
Kennedy, James E. 260, 261, 264, 272, 273, 288, 290, 292, 293, 294, 301, 303, 334, 482
Kennedy, John F. 60
Kennedy, J.L. 280
Kesner, J. 255
Ketcham, K. 60
Khilji, Anjum 338, 366
Kholodov, Y.A. 112
Kilhlstrom, J.F. 68
Kim, C. 386
Kimble 339
Kimura, S. 255
Kirkland, K. 457
Klein, J. 217
Kleinknecht, E.E. 64
Kline, P. 417
Klitch, H. 482
Klyver, N. 218
Knapp, P. 322
Koestler, A. 287
Konig, H. 102
Korinevskaya, I.V. 112
Korinevskii, A.V. 112
Kornbrot, Diana 357, 366
Krippner, Stanley xv, 217, 227, 255, 259, 268, 269, 270, 278, 290, 317, 318, 325, 326, 330, 357, 488
Kristjansson, M. 218
Kross, Ford 335
Krystal, Mr 178
Krystal, Mrs 178
Kugel, W. 219
Kuhn, Robert L. 487
Kuhn, Thomas S. 479, 481, 489
Kulik, J. 60
Kumar, V.K. 68

LaBerge, Stephen 422, 483
Laing, G.P. 436, 438
Lange, R. xv, 99, 100, 102, 111, 113
Laue, M. von 490
Laurence, J.R. 67
Law, E. 100
Lawrence, Tony R. 102, 350, 352, 356, 366
Lemmon, V.W. 290
Lenson, N. 84
Leonard, Gladys O. 205-6 *passim*, 207
Lepper, M.R. 46
LeShan, L. 216, 417
Leuba, L. 290
Levi, A. 273, 289
Levitt, M. 255
Lichtenstein, S. 38, 42
Lieberman, R. 274–5
Liebovici, L. xix
Lindsay, D.S. 60
Lodge, Oliver 160, 162
Loftsson, J. 4
Loftus, E.F. 48, 49, 57, 59, 60, 62, 63, 67, 297
Loftus, G.R. 48, 49
Lombroso, Cesare 160
Lopes, Lola 41, 42, 44, 50
Lord, C. 46
Lowy 285, 286
Lowry, R. 317
Ludwig, L. 218
Lynn, S.J. xv, 61, 62, 488

McAlpine, S. 357
McBain, W. 255, 282, 288, 289, 290
McCaulley, M.H. 338
McClosky, M. 63
McConkey, K.M. 61, 62, 67
McConnell, R.A. 253, 255, 262, 275, 284, 285, 287, 289, 295, 298, 300, 459
McCrae, R.R. 340
McCready, W.C. 4, 25, 30
McCue, P.A. 99
McDermott, K.B. 62, 63, 68, 70
McDonough, Bruce 366
McElroy, B. 386
McFeaters, S.J. 49
MacKenzie, A. 101, 112
McMullin, E. 341
MacRobert, Alan F. 211
McVaugh, M.R. 276, 287
Maher, M.C. xv, 101, 112, 218

Maksovski, T. 69
Markham, R. 67
Marks, David F. xv, 36, 149, 253, 254, 257, 264, 269, 271, 272, 289, 291, 292, 296, 297, 324, 458
Markwick, Betty 157, 256, 299
Martelli, A. 149
Maruri, C.A. 77
Maskelyne, N. 151
Mattuck, R.D. 456
Mauskopf, S.H. 276, 287
May, Edward C. 263, 282, 456, 483, 487
Mayer, B. 255, 270, 274, 291, 294
Mayne, A. 216
Mazzoni, G.A.L. 63
Medhurst, R.G. 299
Meehan, T.C. 436
Mefferd, R.G. 417
Mellen, R.R. 255, 265, 266, 301
Menzies, Stuart 366
Mesmer, Franz A. 454
Metzger, W. 330
Meyers, Vincent 91–7
Michaelson, J.L. 243
Millar, B. 220
Miller, R.N. 435, 437
Mills, A. 84
Milton, Julie xvii, 345–69
Mirabelli, Carlos 167
Mitchell, Edgar D. 480, 484
Molewijk, G.C. 220
Monnet, M. 327, 329, 330
Moody, R.A. xv
Moore, D. 485
Moore, L. 459
Moretti, P. 68, 69
Morris, Robert L. 36, 59, 217, 253, 255, 284, 287, 288, 290, 296, 297, 299, 332, 353, 356, 357, 365, 473, 480, 481, 488
Morse, M. 95
Morselli, Enrico 161, 162
Moss, S.P. 255, 268, 278, 287, 292, 301, 302
Moss, Thelma 217, 253, 255, 268, 270, 277, 285, 287, 288, 290
Mosteller, Frederick 35, 36, 37, 38, 39, 321
Munro, C. 113
Münsterberg, Hugo 165, 166
Murphey, H. 174
Murphy, G. 219, 262, 280, 284
Murray, D.J. 42
Murray, D.M. 44
Musso, J.R. 255, 289, 297, 298
Myers, Frederic 160, 161
Myers, I.B. 338, 408
Myers, S.A. 69

Nader, K. 59, 60
Nadon, R. 68
Nakanishi, M. 255
Nash, C.B. 255, 265, 267, 275
Nash, C.S. 255, 265, 267, 275
Nash, M. 67
Neher, Andrew 324, 458
Neisser, U. 60
Nelson, K. 77
Nelson, Roger D. xviii, 349, 350, 352, 355, 366, 385–400, 485
Neu, J.G. 219, 220
Neuhausler, Dr 213
Newton, Isaac 445
Nickel, J. 102
Nicol, J.F. 255, 256, 266, 275, 277, 298
Nie, N.H. 7
Nisbett, R.E. 39, 45, 50, 329
Nowack, E. 89

Ochorowicz, Julijan 160
Ockham, William of 450, 451
Ofshe, R. 60
O'Keeffe, Ciarán 99–115
O'Mathuna, D. 438, 439
O'Neil, V. 141
O'Rahilly, R. 86
Osis, Karlis xvi, 149, 150, 173–83, 215
Owen, A.RG. 140, 141, 143, 173
Ownbey, Sara *see* Zirkle, Mrs George

Pack, G.T. 77, 78, 84
Paivio, A. 416
Palladino, Eusapia xvi, 157, 158–63 *passim*, 164, 165, 166, 167, 169
Pallikari, Fotini 456, 457, 486
Palmer, John xiv, xvii, xviii, 3–33, 68, 69, 253, 254, 255, 257, 260, 274–5, 288, 294, 300, 330, 334, 340, 355, 356, 361, 366, 422, 430, 447, 452, 453
Panati, C. 276
Parker, Adrian 255, 257, 265, 266, 288, 290, 292, 300, 330, 331, 365, 366, 447, 450, 451, 453, 459, 460
Parnia, S. xv
Parsons, D. 211
Pasricha, S. 78
Paulson, M. 255
Pearce, Hubert 264
Peel, John 121
Pegram, Margaret H. 243

Pekala, R.J. 68, 69
Pembroke, A.C. 89
Perizonius, R. 219
Perry, C. 67
Perry, R.B. 484
Persinger, M.A. 68, 102, 113
Peters, Douglas P. 46
Peters, Ellis 133
Peterson, D.M. 402, 422
Philip, Prince 119
Phillips, L.D. 42
Phillips, P.R. 149, 276
Pickrell, J.E. 63, 67
Piper, Leonora 202
Pirmasens 219
Planck, Max 385, 490
Platt, R.D. 63, 68
Podmore, Frank 165, 169, 202
Polidoro, M. xvi
Poortman, J.J. 422
Pope, D. 293
Pope, K.S. 62
Postman, L.J. 50
Powell, K.F. 262
Powell, R. 102
Prasad, J. 4, 330
Pratt, J.G. xviii, 149, 206–7 *passim*, 211, 215, 216, 217, 218, 223, 243, 249, 256, 264, 267, 281, 287, 288, 291, 293, 296, 305, 341, 354, 447
Price, D. 152
Price, G.R. 265, 272, 298, 299
Puthoff, H. 149, 257, 269, 271, 272, 273, 291, 292, 296, 297, 325, 453, 480
Pynoos, R.S. 59, 60

Quinn, J.F. 436, 437, 438

Radin, Dean I. xviii, 102, 317, 349, 350, 352, 355, 357, 366, 385–400, 478, 486
Ragnarsson, A. 4
Randall, J.L. 149, 151
Randi, J. xv, 149, 150, 253, 276
Rao, K. Ramakrishna xviii, 253, 255, 257, 265, 266, 271, 275, 280, 281, 286, 296, 297, 298, 300, 301, 304, 330, 340, 481, 482
Rao, P.V.K. 255, 257, 265, 266, 271, 280, 281, 297, 298, 302
Rattet, S.L. 69
Ratzenberg, F.H. 68
Raye, C. 64
Read, J.D. 60
Rechtschaffen, A. 322
Redington, D.J. 260
Reinsel, Ruth 487
Reiser, M. 218, 222
Reynolds, C.B. 151
Rhine, Joseph B. xiii, xvii, xviii, 151, 157, 201, 206, 235–51, 256, 262, 264, 267, 276, 281, 287, 290, 291, 293, 295, 296, 298, 299, 305, 330, 341, 451, 452, 453, 455, 481
Rhine, Louisa E. xviii, 3, 14, 330, 481
Richards, A. 57
Richards, D.G. 68, 69
Richet, Charles 159, 160, 173, 239
Rinaldi, G.M. xvi
Ring, K. xv, 95
Roberts, J.A. 102
Roberts, L. xix
Roberts, R. xv
Roberts, R.R. 331
Robson, C. 425
Roe, Chris xiv, 366, 448, 453
Roediger, H.L. 62, 63, 68, 70
Roentgen, Wilhelm C. 452
Rogo, D. Scott 166, 278
Roig, M. 330, 487
Roll, M. 217
Roll, W.G. 102, 140, 141, 142, 143, 144, 149, 173, 215, 216, 217, 223, 418
Romm, Ethel G. 324, 325
Rosa, L. 434
Rose, N. 64, 70
Rosenthal, R. 105, 269, 280, 282, 298, 321, 332, 334, 338, 339, 352, 396, 414, 415
Rosnow, R.L. 105, 339
Ross, C.A. 69
Ross, L. 39, 45, 46, 329
Rosza, Mikla 326
Roth, J.A. 299
Roy, D.F. 50
Rubin, D.B. 269
Rudolph, L. 284
Ruíz-Noguez, L. 141
Russell, D. 59
Ruthchild, M. 149
Rutowski, C.A. 102
Ryzl, M. 255, 267, 286, 287

Sabnani, D. 177
Sabom, M.B. 95
Sadler, T.G. 417
Sai Baba, Sathya xvi, 174–82 *passim*
Saltmarsh, H.F. 203–6 *passim*, 207, 226, 239, 240
Samararatne, G. 78
Samuels, M. 490
Samuels, N. 490

Sandweiss, S.H. 174
Sannwald, G. 330
Sargent, Carl L. 255, 266, 268, 273, 277, 278, 281, 282, 286, 287, 288, 294, 301, 302, 468, 469, 471, 482
Saunders, D.R. 332
Saxe, S. 218
Schacter, D.L. 62
Schechter, Ephraim I. 330, 338
Schlitz, Marilyn J. xix, xx, 336, 340, 366, 416, 421–31, 437, 477–92
Schmeidler, Gertrude R. xv, 102, 112, 216, 217, 218, 255, 259, 264, 265, 266, 269, 275, 278, 279, 280, 284–5 *passim*, 287, 289, 295, 298, 336, 361
Schmidt, Helmut xviii, 259, 260, 262, 263, 283, 283–5 *passim*, 299, 340, 377–83, 456, 471, 472, 482
Schmidt, Stefan 486
Schmitt, M. 255, 265, 266, 274
Schneider, Rudi 174
Schouten, Sybo A. xiv, xvi, 199–232, 256, 300
Schrenck-Notzing, A. von 173
Schrödinger, Erwin 385
Schulman, A. 174
Scott, Christopher xvii, 156, 208, 209, 210, 256, 280, 284–5 *passim*, 287, 289, 290, 299
Sells, S.B. 293
Serios, Ted 174
Shafer, Donna 401–19, 421, 422
Shafer, M. 149, 276
Shapiro, H.A. 82
Sheehan, P.W. 67
Sheehy, Gail 193
Sheldrake, Rupert 486
Sherman, Steven J. 43, 44, 453
Shermer, M. xv
Sherwood, Simon J. 448
Shimony, A. 386
Shields, E. 255, 257, 258, 265, 266, 267, 269, 270, 271, 275, 278, 279, 280, 281, 286, 287, 297, 298
Shrager, E.F. 255
Sicher, F. 435, 437, 438, 485
Sidgwick, Eleanor 160, 161, 164
Sidgwick, Henry 4, 13, 153, 160, 161
Siegel, Robert K. 474
Silva, Fabio da 366
Simington, J.A. 436, 438
Singer, B. 151
Skinner, J. 322
Slovic, P. 38, 39
Smith, B.M. 256
Smith, D. 357
Smith, D.W. 78
Smith, E.E. 330, 357
Smith, H.S. 485
Smith, Matthew D. 366, 422, 450
Snowdon, R.J. 262
Soal, S.G. 204–6 *passim*, 207, 256, 299, 453
Solley, Charlie 480
Sondow, N. 255, 257, 268, 273, 274, 277, 278
Spalteholz, W. 84
Spanos, N.P. 61, 65, 66, 68, 69
Spence, K.W. 339
Spinelli, E. 468
Spottiswoode, S.J.P. 456, 486
Standish, Leanna 489
Stanford, Rex G. 255, 260, 262, 265, 266, 270, 274, 276, 291, 294, 302, 340, 350, 352, 354, 355, 356, 456
Stapp, Henry 486
Stebbins, R.A. 149
Stein, A.G. 350, 352, 354, 355, 356
Stein, S. 261
Steinbrenner, K. 8
Steinkamp, Fiona 350, 353, 366
Stendhal 89
Stepanek 254
Sterling, T.C. 331
Stern, L.D. 49
Stevens, Paul 99–115
Stevenson, Ian xv, 4, 77–90, 216, 218, 330
Stokes, D.M. 257, 269, 289, 341
Storm, L. xvii
Strauch, I. 322
Stuart, C.E. 256, 273
Stump, J. 255, 257, 265, 266, 270, 290, 291, 301
Sundberg, N.D. 187
Sweeney, V.M. 322
Swets, J.A. 334
Symmons, Charles 365, 366

Taddonio, J.L. 288
Taft, R. 68
Tandy, V. 102
Targ, Elisabeth 485, 487, 488
Targ, R. 149, 257, 269, 271, 272, 273, 291, 292, 296, 297, 325, 438, 439, 453, 480
Tart, Charles T. 215, 260, 261, 262, 273, 287, 460, 473
Taylor, Kirsten I. 459, 460
Taylor, R. 357
Taylor, S.E. 41, 44
Teguis, A. 141
Tellegen, A. 67
Tenhaeff, W.H.C. 219
Teresa of Calcutta, Mother 119

Terry, James C. 255, 257, 268, 270, 271, 274, 277, 278, 279, 289, 292, 294, 366
Thomas, C. Drayton 207
Thomas, J.F. 205–6 *passim*
Thouless, R.H. 211
Thurley, S. 100
Thurston, Howard 149
Tiller, S.G. 113
Timm, Ulrich 214, 215, 217, 219
Tirado, J. 255
Titchener, E.B. 422
Tobacyk 70
Tomes, J.L. 67
Tort, C.J. 141
Troscianko, T.S. 41, 468
Truzzi, M. xvi, 149, 253
Turner, J.G. 436
Tversky, Amos 39, 40, 41, 43, 50, 329, 338, 339
Twain, Mark 475
Tyrrell, G.N.M. 212, 217

Ullman, Montague 217, 227, 269, 270, 317, 318, 319, 320, 325, 326, 330
Underwood, P. 100, 102
Uphoff, H.F. 280
Utts, Jessica xvii, 330, 332, 333, 339, 349, 366, 456

Van de Castle, Robert L. 322, 323
Van Der Velden, I. 68, 69
Vandrey, R. 149
Van Gogh, Vincent 325, 326
van Lommel, Pim xv, 69, 91–7
van Wees, Roud 91–7
Varvoglis, Mario 366, 486
Vassar, C. 274, 294
Vaughan, A. 217, 270, 317, 330
Vett, Carl 179
Voltaire 155
Von Lucadou, Walter 486
Voois, W. 220
Voustianiouk, A. 112

Wagner, C. 218
Wagner, M.W. 68, 327, 329, 330
Wagstaff, G.F. 61
Walach, H. 486
Walker, E.H. 456
Walker, S.R. 435, 438
Walter, R.D. 102
Warren, Charles 366
Wasserman, T. 422
Watt, Caroline A. xv, 35–53, 99–115, 366
Waters, E. 60
Watkins, Anita 481
Watkins, Graham K. 481
Watson, D. 407, 416
Weiner, Debbi 482
Weiner, H. 261
Welldon, R.M. 435
Wells, G.L. 44, 48
Wendlandt, S. 149
West, D.J. 212
Westerlund, Joakin 365, 366
Wezelman, Rens 365, 366
White, R.A. 288
Wicklund, N. 293
Wickramasekera, I. 68
Wiklund, Nils 366
Wilbur, Ken 484
Wilkinson, H.P. 102
Williams, Carl 100, 366
Williams, H.C. 89
Williams, L.M. 61, 402
Willin, Melvyn 366
Wilson, A.J. 100
Wilson, J.G. 78, 267
Wilson, R. 256, 261
Wilson, W.R. 255, 265, 270, 271, 277, 279
Win Maung, G. 78
Windle, C. 275
Winifred, St 132, 133
Winograd, E. 68
Wirth, D.P. 436, 437
Wiseman, Richard xv, xvi, xvii, xix, 58, 59, 99–115, 347, 348, 351, 353, 356, 357, 358, 359, 361, 366, 421–31, 461, 483
Wolfradt, U. 69
Wolman, B.B. 253, 254, 276, 322
Wood, R. 255, 268, 274, 277, 278, 292, 294, 330
Woodruff, J.L. 264, 296

Yapko, M.D. 60–61
Yogananda, P. 173
Yount, Garret 485, 489

Zaragoza, M.S. 63
Zirkle, George 243
Zirkle, Mrs George 243, 249
Zusne, Leonard xv, 253, 258, 322, 325–6